QTL Studies- Past, Present and Future

David Evans

University of BRISTOL

Genetic studies of complex diseases have not met anticipated success

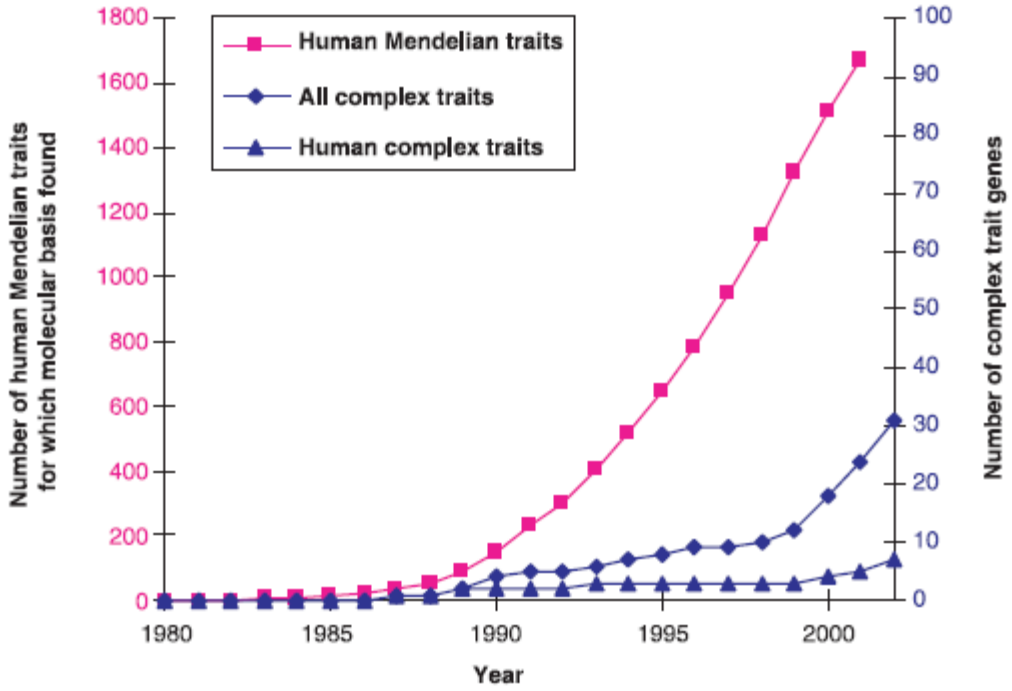


Fig. 1. Identification of genes underlying human Mendelian traits and genetically complex traits in humans and other species. Cumulative data for human Mendelian trait genes (to 2001) include all major genes causing a Mendelian disorder in which causal variants have been identified (*58*, *59*). This reflects mutations in a total of 1336 genes. Complex trait genes were identified by the whole-genome screen approach and denote cumulative year-on-year data described in this review.

Glazier et al, Science (2002) 298:2345-2349

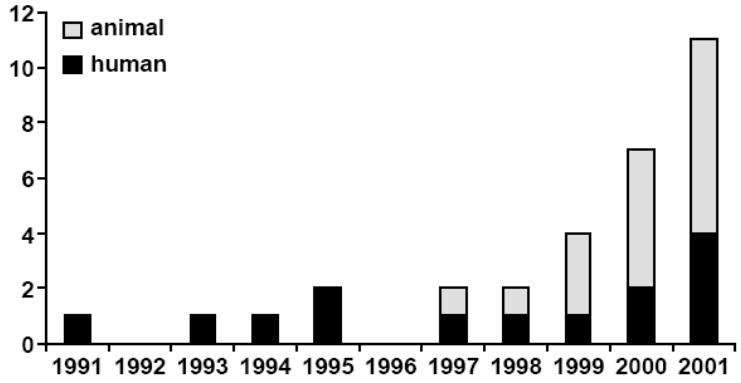


Fig. 1 Number of genes identified from QTL by year. Genes for human QTL are shown in black and genes for experimental models (mouse, rat, and pig) in white. The first QTL gene was identified in 1991. There are 14 from humans and 17 from animal models (5 from rats, 11 from mice, 1 from pigs). These add up to more than 29 because some were identified in both humans and rodent models.

Korstanje & Pagan (2002) *Nat Genet*

Table 1 • Genes identified from QTL studies

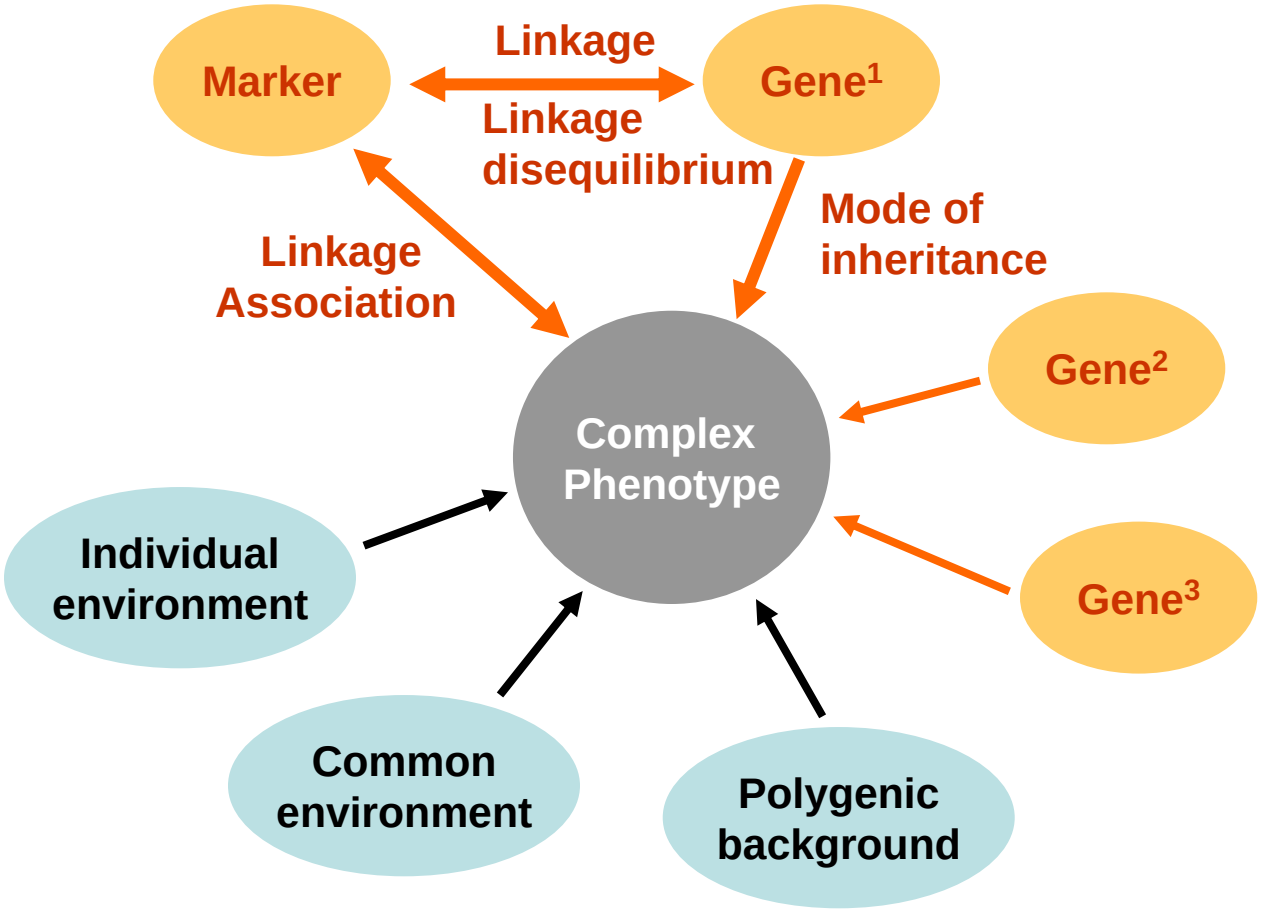
Polygenic trait	Year	Ref.	Gene	Species	pos	tg	ko	fu
Alzheimer disease	1991	9	*APP*	human				X
Alzheimer disease	1993	10	*APOE*	human				
Ovarian and breast cancer	1994	11	*BRCA1*	human	X			X
Breast cancer	1995	12	*BRCA2*	human	X			X
Insulin resistance	1995	13	*FABP2*	human				
HDL-cholesterol levels	1997	14	*LIPC*	human				
Intestinal cancer	1997	15	*Pla2g2a*	mouse	X	X		
Blood pressure	1998	16	*Atp1a1 / ATP1A1*	rat/human		X		X
Leptin levels	1999	17,18	*POMC*	human			X[a]	X
Asthma	1999	19	*Il4*	mouse	X	X		
Asthma	1999	19	*Il13*	mouse	X	X		
Insulin-mediated glucose uptake	1999	2	*Cd36*	rat		X		
Obesity	2000	20	*Ptpn1/PTPN1*	mouse/human			X[b]	X
Alzheimer disease	2000	21	*PSEN1*	human	X			
Diabetes	2000	22	*Il2*	mouse	X		X[b]	X
Gallstones	2000	23	*Abcc2*	mouse	X			X
Asthma	2000	3	*Hc*	mouse				
Muscle glycogen content	2000	24	*Prkag3*	pig	X		X[c]	X
Crohn disease	2001	25,26	*NOD2*	human	X		X[a]	X
Blood pressure	2001	27	*SCNN1A1*	human			X[a]	
Blood pressure	2001	28	*SCNN1G*	human			X[a]	
Blood pressure	2001	29	*Slc12a1*	rat				
Blood pressure	2001	30	*Cyp11b1*	rat				X
Bone density	2001	5	*COL1A*	human				
Left ventricular mass	2001	31	*Nppa*	rat			X[b]	X
Modifier of tubby hearing	2001	32	*Mtap1a*	mouse	X	X		X
Taste, saccharin response	2001	33	*Tas1r3*	mouse	X	X		X
Tumor susceptibility	2001	34	*Cdkn2a*	mouse	X		X[b]	X
Diabetes	2001	35	*B2m*	mouse		X	X	

pos, found by positional cloning; tg, transgenic insertion of normal gene changes phenotype to normal (for example, transgenic rescue); ko, knockout provides additional evidence ([a]human monogenic syndrome, [b]deletion of gene by homologous recombination produces a mouse with the phenotype typical of the disease, [c]knockout in yeast); fu, functional difference in candidate gene. *APP*, amyloid precursor protein; *APOE*, apolipoprotein E; *BRCA*, breast cancer gene; *FABP2*, fatty acid binding protein 2; *LIPC*, hepatic lipase; *ATP1A1*, α-Na,K-ATPase; *POMC*, pre-pro-opiomelanocortin; *Il*, interleukin; *Cd36*, fatty acid translocase; *PTP1B*, protein tyrosine phosphatase-1B; *PSEN*, presenilin 1; *Abcc2*, ATP-binding cassette, subfamily C2; *Hc*, hemolytic complement (*C5*); *Prkag3*, protein kinase, AMP-activated, γ3; *NOD2*, caspase recruitment domain-containing protein 15 (*CARD15*); *SCNN*, sodium channel, non-voltage gated; *Slc12a1*, Na,K,2Cl-cotransporter; *Cyp11b1*, 11β-hydroxylase; *COL1A*, collagen-1A; *Nppa*, natriuretic peptide precursor A; *Mtap1a*, microtubule-associated protein 1a; *Tas1r3*, taste receptor-3; *Cdkn2a*, cyclin-dependent kinase inhibitor 2a; *B2m*, β2-microglobulin.

Korstanje & Pagan (2002) *Nat Genet*

Reasons for Failure?

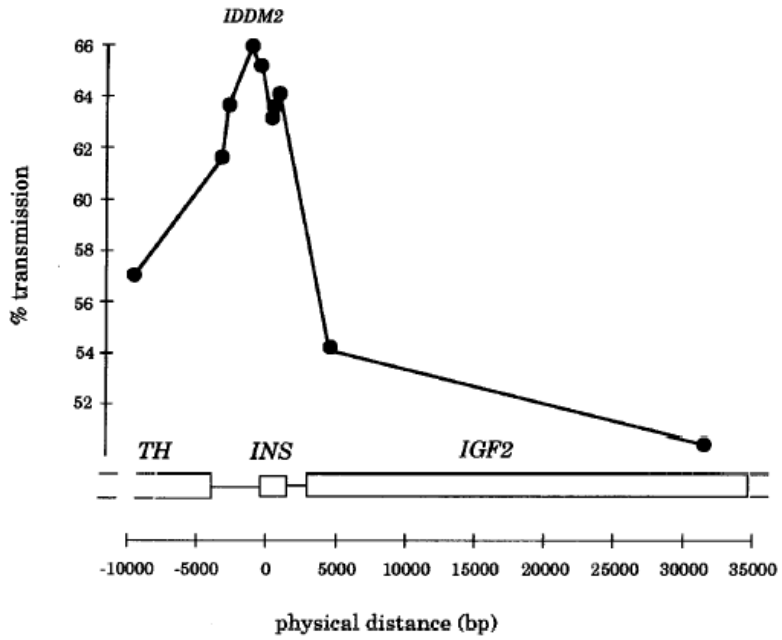
- BUT…Not much success in mapping complex diseases / traits !



Weiss & Terwilliger (2000) *Nat Genet*

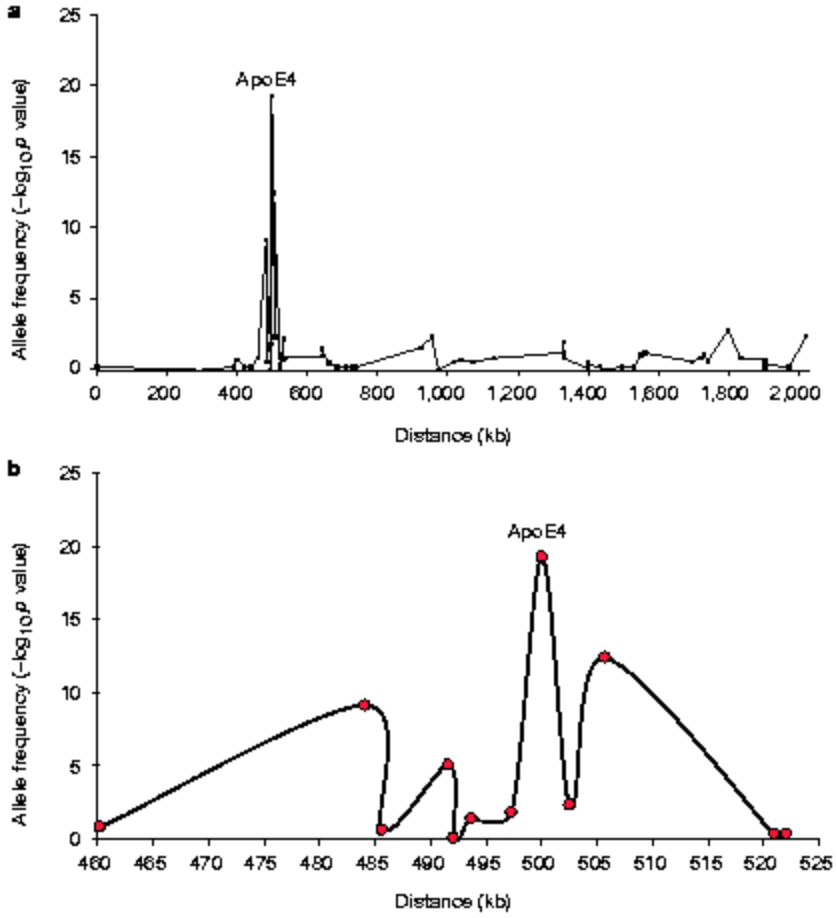
LD Patterns and Allelic Association

Type 1 diabetes and Insulin VNTR



Bennett & Todd, *Ann Rev Genet*, 1996

Alzheimers and ApoE4



Roses, *Nature* 2000

- Pattern of LD unpredictable
- Extent of common genetic variation unknown

Genome-wide Association?

The Future of Genetic Studies of Complex Human Diseases

Neil Risch and Kathleen Merikangas

		Linkage			Association			
					Singletons		Sib pairs	
Genotypic risk ratio (γ)	Frequency of disease allele A (p)	Probability of allele sharing (Y)	No. of families required (N)	Probability of transmitting disease allele A P(tr-A)	Proportion of heterozygous parents (Het)	(N)	(Het)	(N)
4.0	0.01	0.520	4260	0.800	0.048	1098	0.112	235
	0.10	0.597	185	0.800	0.346	150	0.537	48
	0.50	0.576	297	0.800	0.500	103	0.424	61
	0.80	0.529	2013	0.800	0.235	222	0.163	161
2.0	0.01	0.502	296,710	0.667	0.029	5823	0.043	1970
	0.10	0.518	5382	0.667	0.245	695	0.323	264
	0.50	0.526	2498	0.667	0.500	340	0.474	180
	0.80	0.512	11,917	0.667	0.267	640	0.217	394
1.5	0.01	0.501	4,620,807	0.600	0.025	19,320	0.031	7776
	0.10	0.505	67,816	0.600	0.197	2218	0.253	941
	0.50	0.510	17,997	0.600	0.500	949	0.490	484
	0.80	0.505	67,816	0.600	0.286	1663	0.253	941

Comparison of linkage and association studies. Number of families needed for identification of a disease gene.

Risch & Merikangas, *Science* 1996

Multiple Rare Variant Hypothesis?

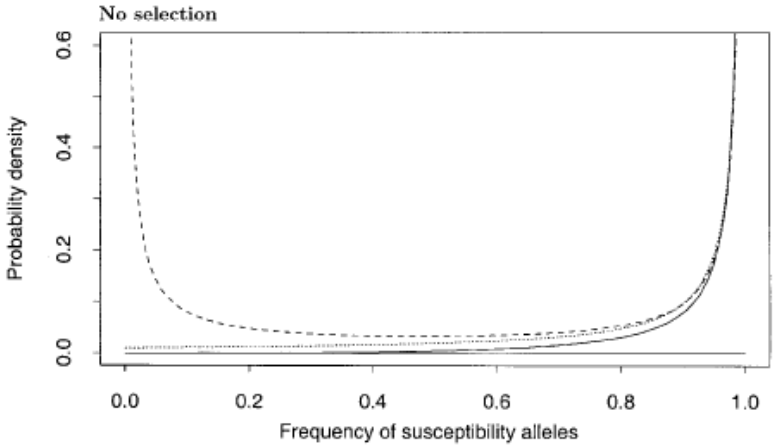
- GWA assumes that common variants underlie common diseases

Are Rare Variants Responsible for Susceptibility to Complex Diseases?

Jonathan K. Pritchard

Department of Statistics, University of Oxford, Oxford

Little is known about the nature of genetic variation underlying complex diseases in humans. One popular view proposes that mapping efforts should focus on identification of susceptibility mutations that are relatively old and at high frequency. It is generally assumed—at least for modeling purposes—that selection against complex disease mutations is so weak that it can be ignored. In this article, I propose an explicit model for the evolution of complex disease loci, incorporating mutation, random genetic drift, and the possibility of purifying selection against susceptibility mutations. I show that, for the most plausible range of mutation rates, neutral susceptibility alleles are unlikely to be at intermediate frequencies and contribute little to the overall genetic variance for the disease. Instead, it seems likely that the bulk of genetic variance underlying diseases is due to loci where susceptibility mutations are mildly deleterious and where there is a high overall mutation rate to the susceptible class. At such loci, the total frequency of susceptibility mutations may be quite high, but there is likely to be extensive allelic heterogeneity at many of these loci. I discuss some practical implications of these results for gene mapping efforts.



How many diseases does it take to map a gene with SNPs?

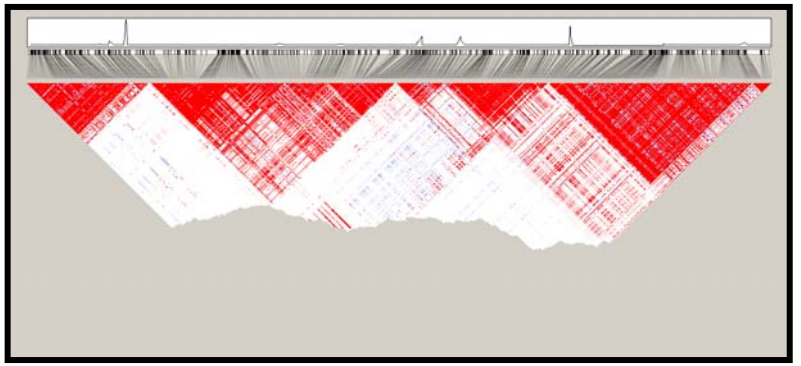
Kenneth M. Weiss[1] & Joseph D. Terwilliger[2]

"They all talked at once, their voices insistent and contradictory and impatient, making of unreality a possibility, then a probability, then an incontrovertible fact, as people will when their desires become words."
—W. Faulkner, *The Sound and the Fury*, 1929

"I found one! I found one!"

Enabling Genome-wide Association Studies

- HAPlotype MAP



- High throughput genotyping

- Large cohorts

CHILDREN OF THE 90s

"ALSPAC"

AUSTRALIAN twin REGISTRY

VRIJE UNIVERSITEIT NEDERLANDS TWEELINGEN REGISTER

Wellcome Trust Case Control Consortium

Wellcome Trust Case-Control Consortium

Genome-Wide Association Across Major Human Diseases

DESIGN

Collaboration amongst 26 UK disease investigators
2000 cases each from 9 diseases
1000 cases from 4 diseases

GENOTYPING

Affymetrix 500k SNPs
Illumina Human NS_12 SNP chip

CASES

1. Type 1 Diabetes
2. Type 2 Diabetes
3. Crohn's Disease
4. Coronary Heart Disease
5. Hypertension
6. Bipolar Disorder
7. Rheumatoid Arthritis
8. Malaria
9. Tuberculosis

10. Ankylosing Spondylitis
11. Grave's Disease
12. Breast Cancer
13. Multiple Sclerosis

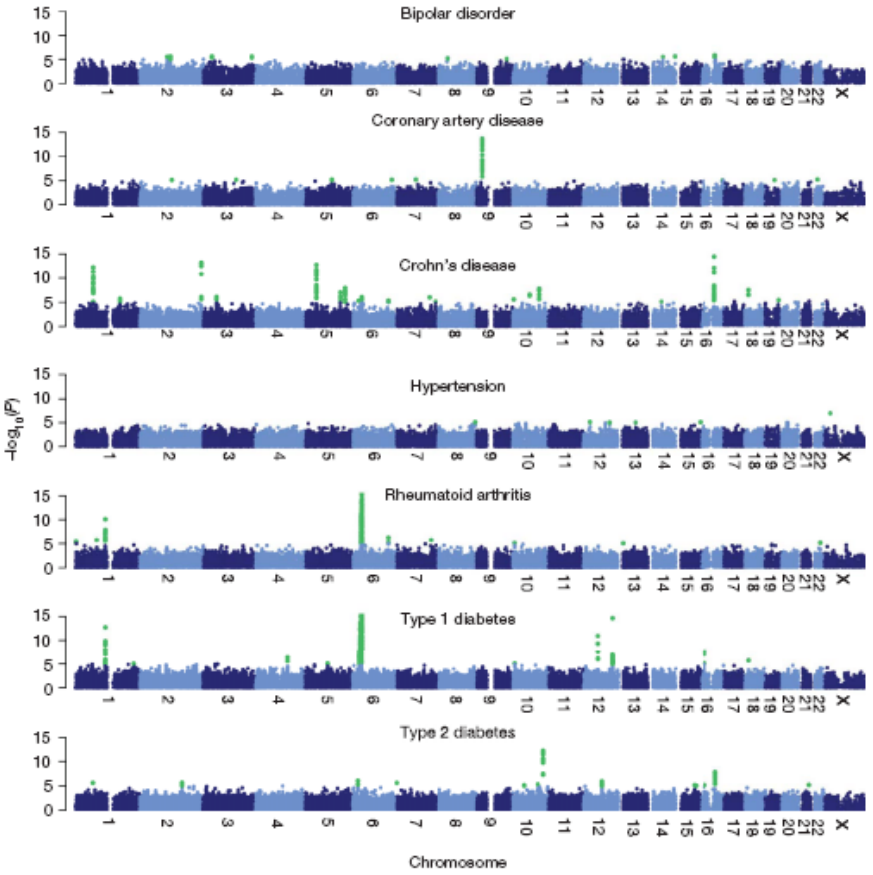
CONTROLS

1. UK Controls A (1,500 - 1958 BC)
2. UK Controls B (1,500 - NBS)
3. Gambian controls (2000)

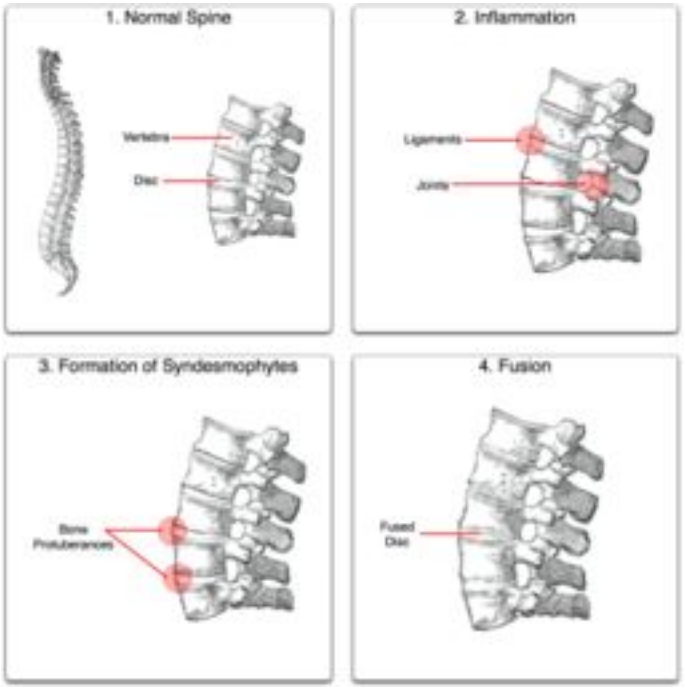
Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

The Wellcome Trust Case Control Consortium*

There is increasing evidence that genome-wide association (GWA) studies represent a powerful approach to the identification of genes involved in common human diseases. We describe a joint GWA study (using the Affymetrix GeneChip 500K Mapping Array Set) undertaken in the British population, which has examined ~2,000 individuals for each of 7 major diseases and a shared set of ~3,000 controls. Case-control comparisons identified 24 independent association signals at $P < 5 \times 10^{-7}$: 1 in bipolar disorder, 1 in coronary artery disease, 9 in Crohn's disease, 3 in rheumatoid arthritis, 7 in type 1 diabetes and 3 in type 2 diabetes. On the basis of prior findings and replication studies thus-far completed, almost all of these signals reflect genuine susceptibility effects. We observed association at many previously identified loci, and found compelling evidence that some loci confer risk for more than one of the diseases studied. Across all diseases, we identified a large number of further signals (including 58 loci with single-point P values between 10^{-5} and 5×10^{-7}) likely to yield additional susceptibility loci. The importance of appropriately large samples was confirmed by the modest effect sizes observed at most loci identified. This study thus represents a thorough validation of the GWA approach. It has also demonstrated that careful use of a shared control group represents a safe and effective approach to GWA analyses of multiple disease phenotypes; has generated a genome-wide genotype database for future studies of common diseases in the British population; and shown that, provided individuals with non-European ancestry are excluded, the extent of population stratification in the British population is generally modest. Our findings offer new avenues for exploring the pathophysiology of these important disorders. We anticipate that our data, results and software, which will be widely available to other investigators, will provide a powerful resource for human genetics research.

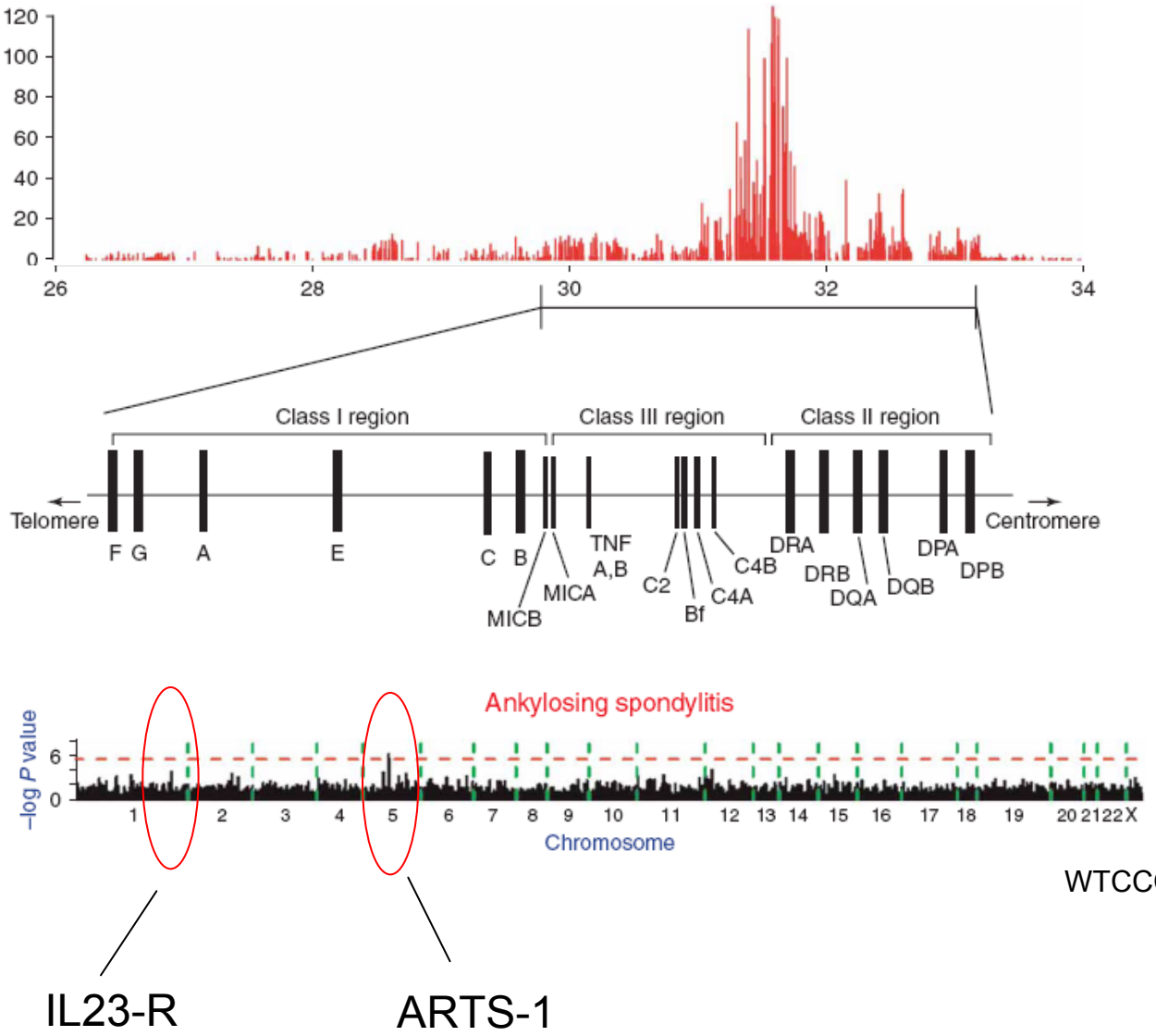


Ankylosing Spondylitis



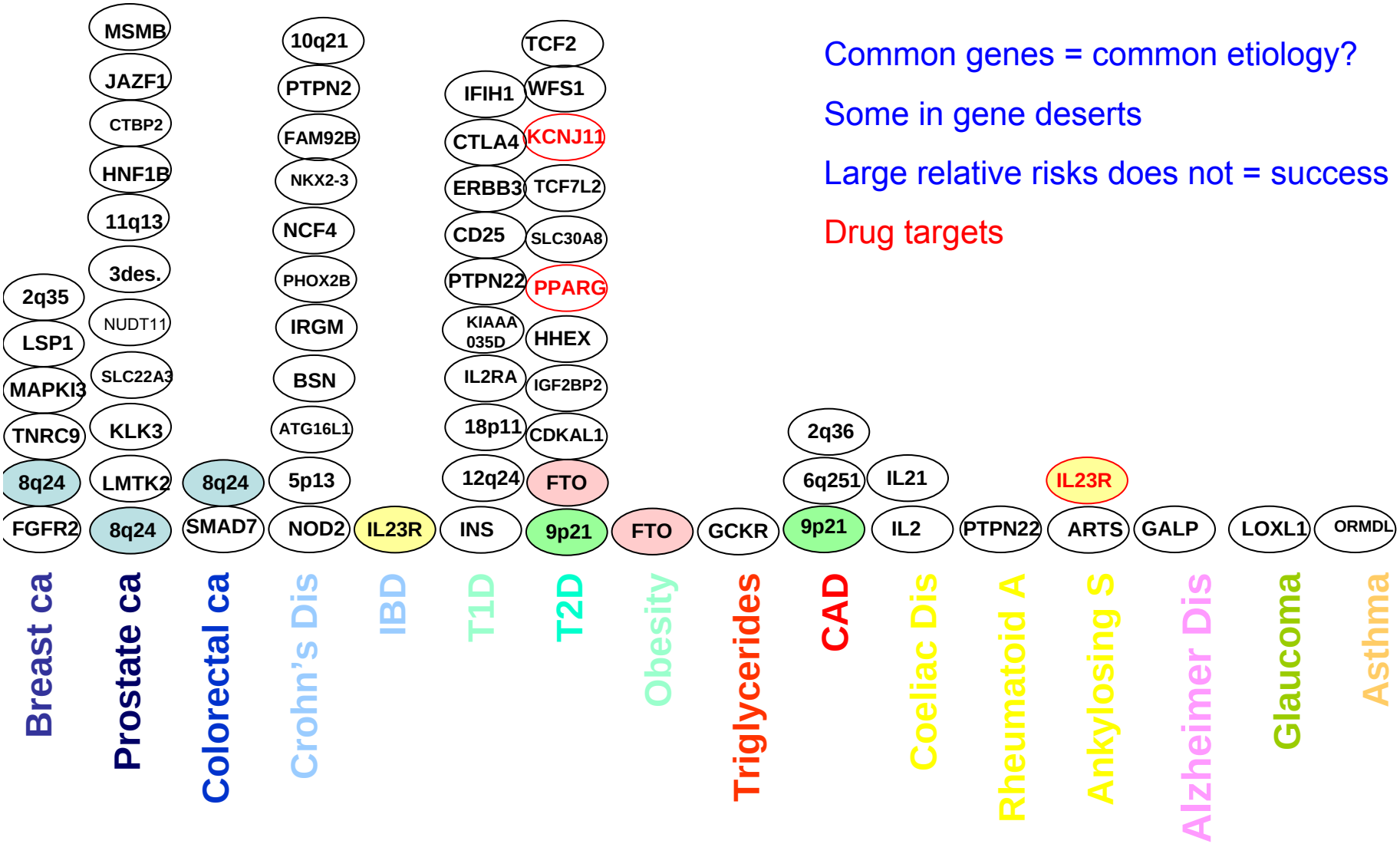
- Auto-immune arthritis resulting in fusion of vertebrae
- Prevalence of 0.4% in Caucasians. More common in men.
- Often associated with psoriasis, IBD and uveitis
- Ed Sullivan, Mike Atherton

Ankylosing Spondylitis GWAS



WTCCC (2007) *Nat Genet*

Successes…



Common genes = common etiology?

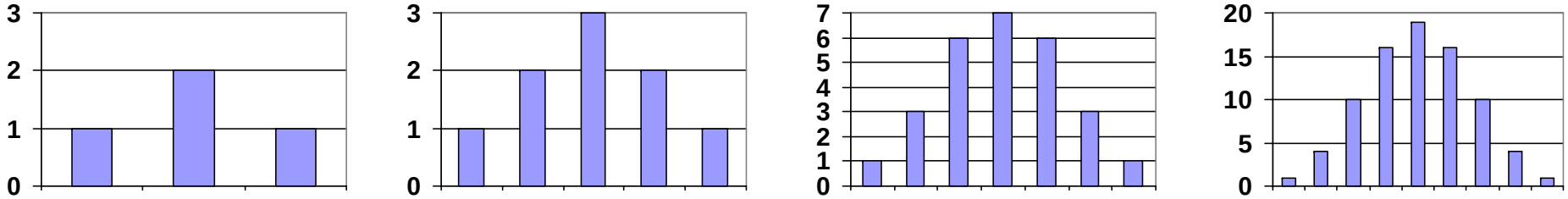
Some in gene deserts

Large relative risks does not = success

Drug targets

What About Quantitative Traits?

1 Gene	2 Genes	3 Genes	4 Genes
→ 3 Genotypes	→ 9 Genotypes	→ 27 Genotypes	→ 81 Genotypes
→ 3 Phenotypes	→ 5 Phenotypes	→ 7 Phenotypes	→ 9 Phenotypes



Central Limit Theorem → Normal Distribution

- Quantitative genetics theory suggests that quantitative traits are the result of many variants of small effect
- Unselected samples
- The corollary is that very large sample sizes will be needed to detect these variants in UNSELECTED samples

Sections
Latest News
Scotland
- Aberdeen
- Dundee
- Edinburgh
- Glasgow
- Inverness
UK
International
Politics
Sci-Tech
Health
Education
Entertainment
Gaelic
Opinion
Obituaries
Article Index
Other Sections
Supplements
Topics
Topics A-Z
Other Sites
The Scotsman

THE SCOTSMAN
Published Date: 13 April 2007
Source: The Scotsman
Location: Scotland
Related Topics: Obesity

Email a friend
Print article
Increase text size
Decrease text size
23 comments on this article

Does my bum look big in these genes?

By LYNDSAY MOSS
HEALTH CORRESPONDENT

OBESITY can now legitimately be blamed on your genes, scientists said yesterday.

Variations in a gene carried by 16 per cent of the population can make people up to 70 per cent more likely to become obese, according to a major study involving thousands of Scottish volunteers.

A genetic variation identified by scientists can make people 70 per cent more likely to be obese. Picture: AFP/Getty Images

The researchers said they had now identified the clearest genetic link yet to weight gain and obesity in the population.

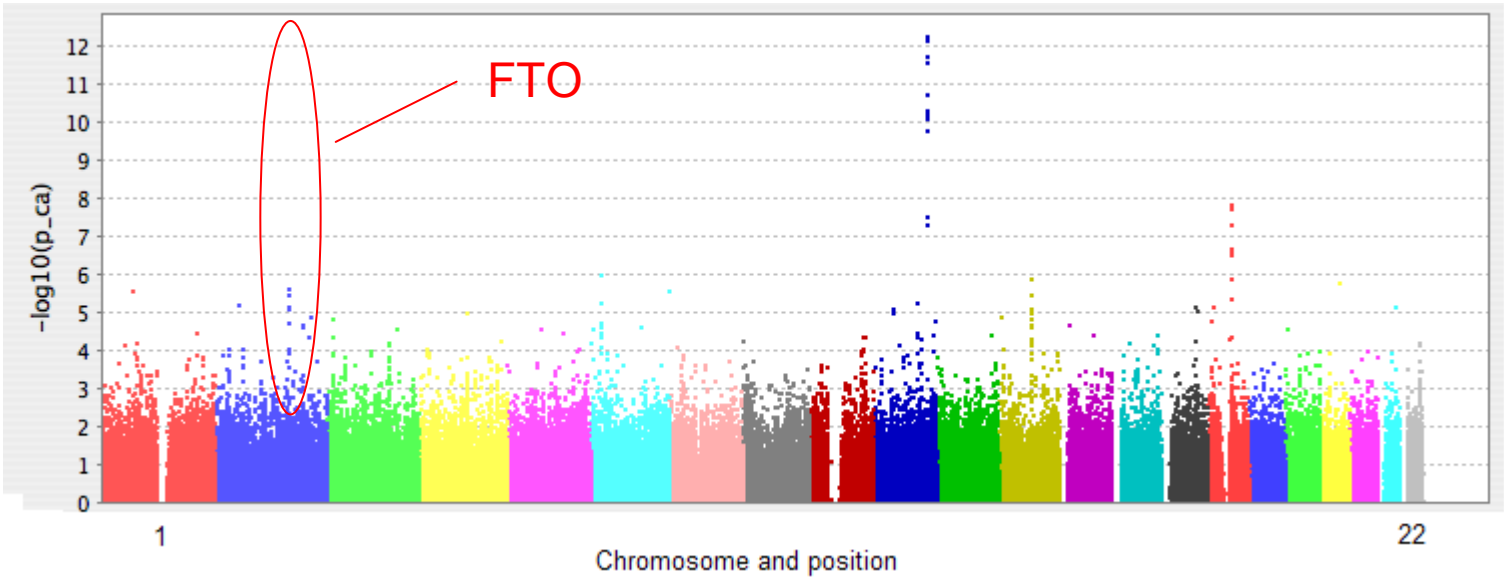
The findings could eventually lead to new treatments for obesity, which affects more than a fifth of adults in the UK.

The research, funded by the Wellcome Trust, studied the genome - all the genetic material in a living object - in 2,000 people with type-2 diabetes and a 3,000-strong control group.

More Obesity:
- Cheese could carry a health alert
- Experts see families as key to beating childhood obesity
- Expert see families as key to beating childhood obesity
- Royal welcome to first 'Trim Town'
- Obese face 50% fat tax for life insurance
- Scotland must play ball to cut obesity
- Women put weight gain down to work
- University seeks to flesh out database
- Poll links obesity to being out of control financially
- Calorie counts for the Big Apple
- Fat's the way to do it for a longer life
- Obese patients in Lothian to be offered stomach surgery
- Only one in three adults gets enough exercise

FTSO

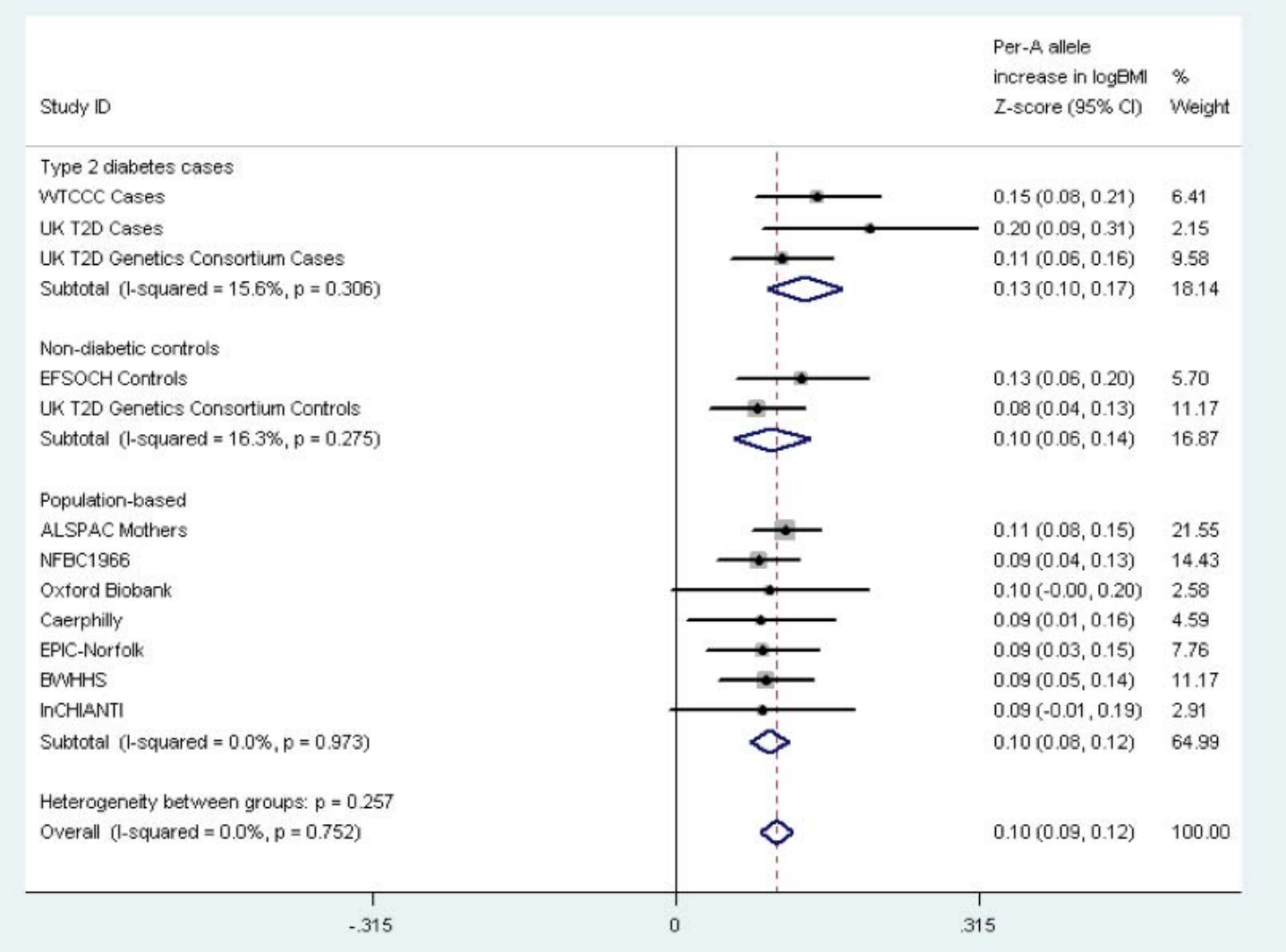
WTCCC T2D Scan



- FTO produces a moderate signal in WTCCC T2D scan
- But, no signal in an American T2D scan…?

  -American cases and controls matched on BMI

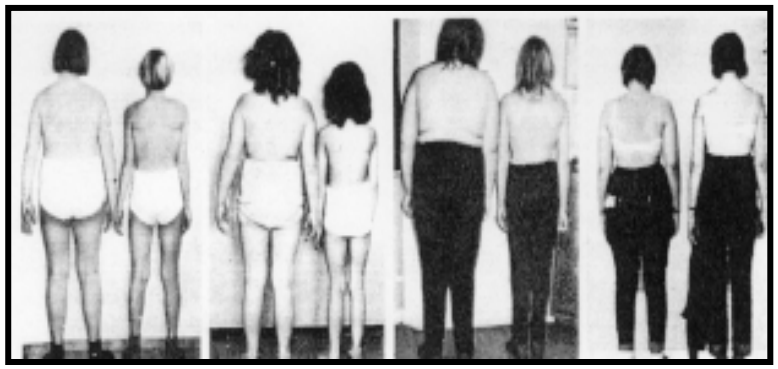
FTO



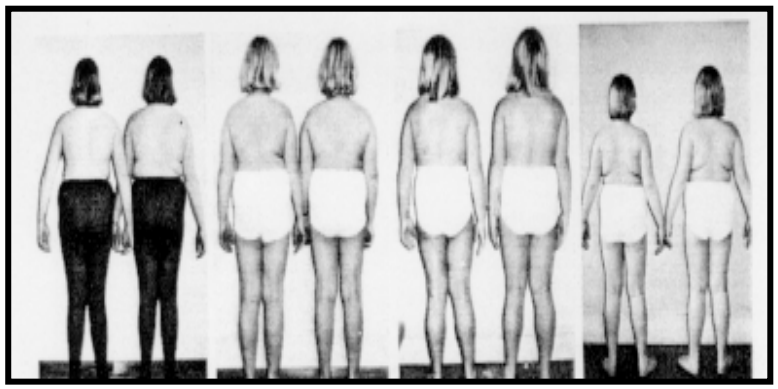
- Replication is critical !!!

Height- The Archetypal Polygenic Trait

Dizygotic Twins



Monozygotic Twins

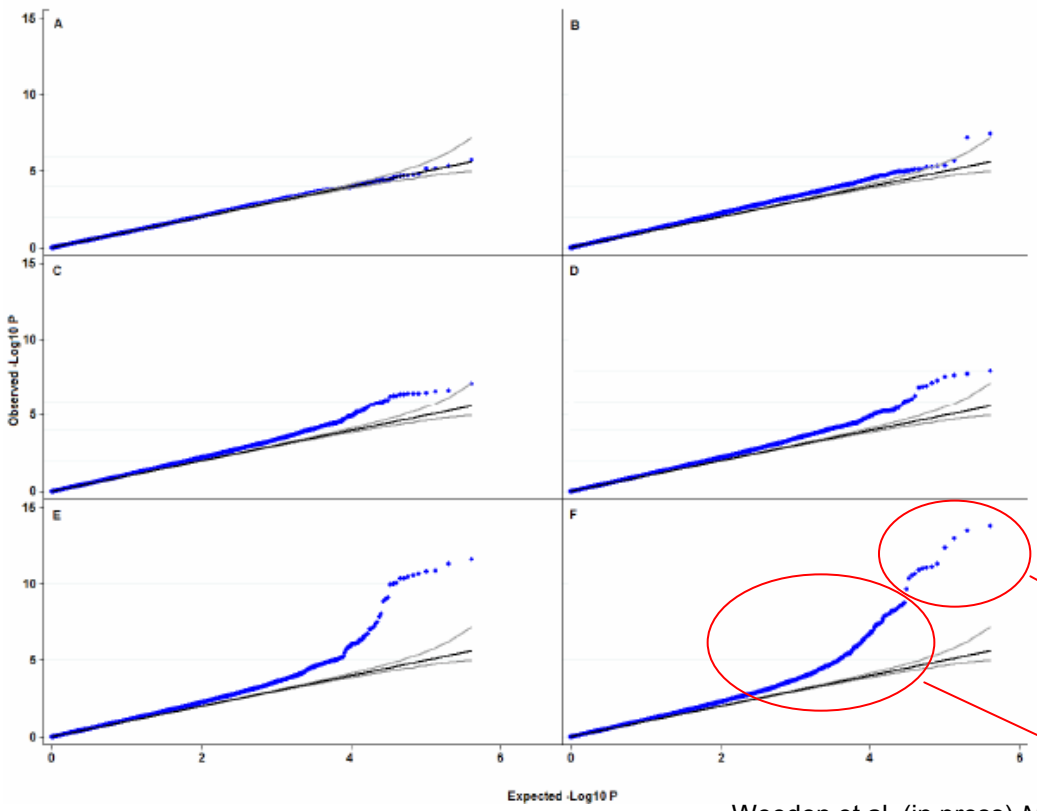


Twins separated at birth

Borjeson, Acta Paed, 1976

Twin, family and adoption studies suggest that, within a population, 90% of variation in height is due to genetic variation

GWA of Height



Weedon et al. (in press) *Nat Genet*

A- 1914 Cases (WTCCC T2D)

B- 4892 Cases (DGI)

C- 6788 Cases (WTCCC HT)

D- 8668 Cases (WTCCC CAD)

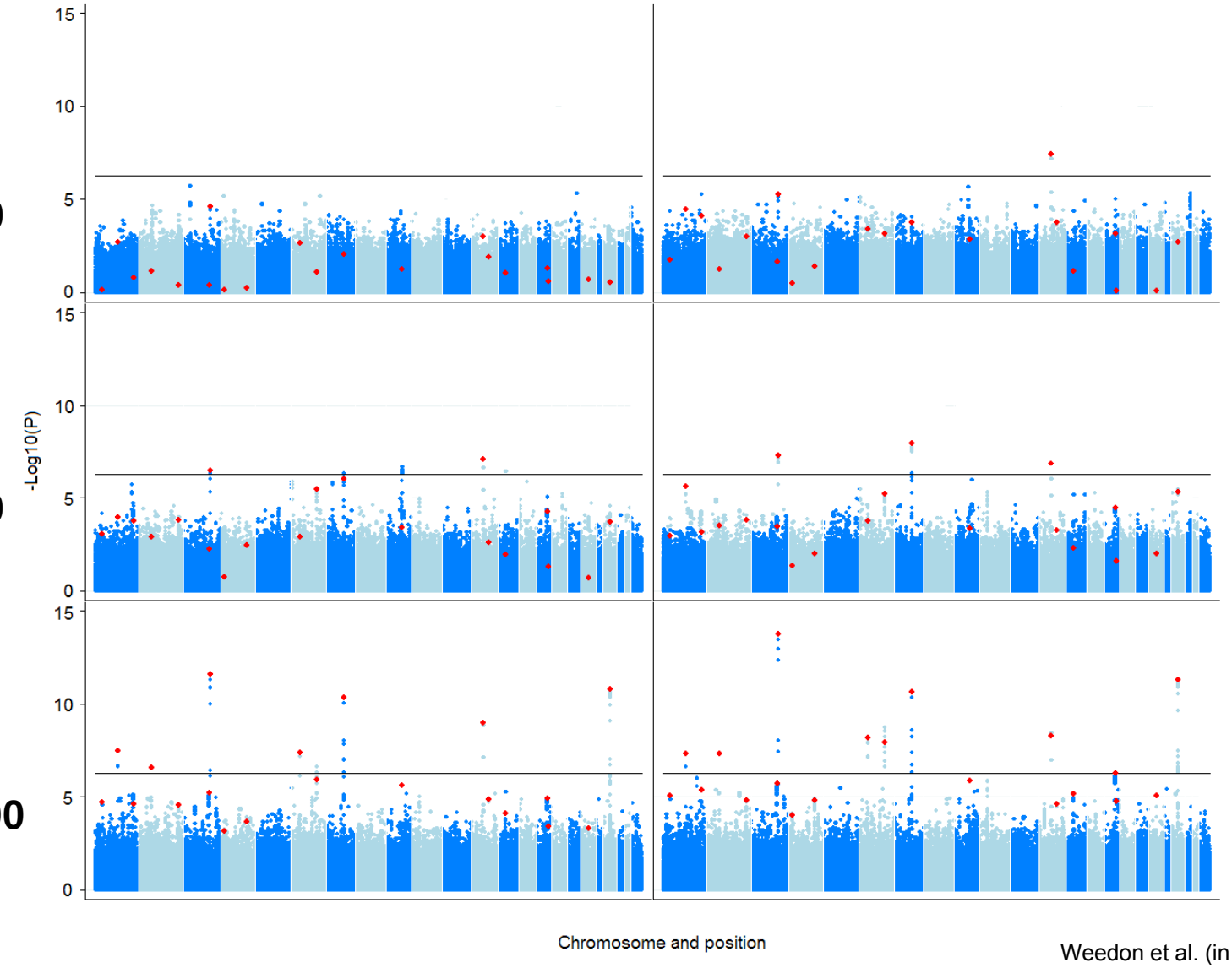
E- 12228 Cases (EPIC)

F- 13665 Cases (WTCCC UKBS)

Significant results

Other loci?

- Large numbers are needed to detect QTLs !!!
- Collaboration is the name of the game !!!



Weedon et al. (in press) *Nat Genet*

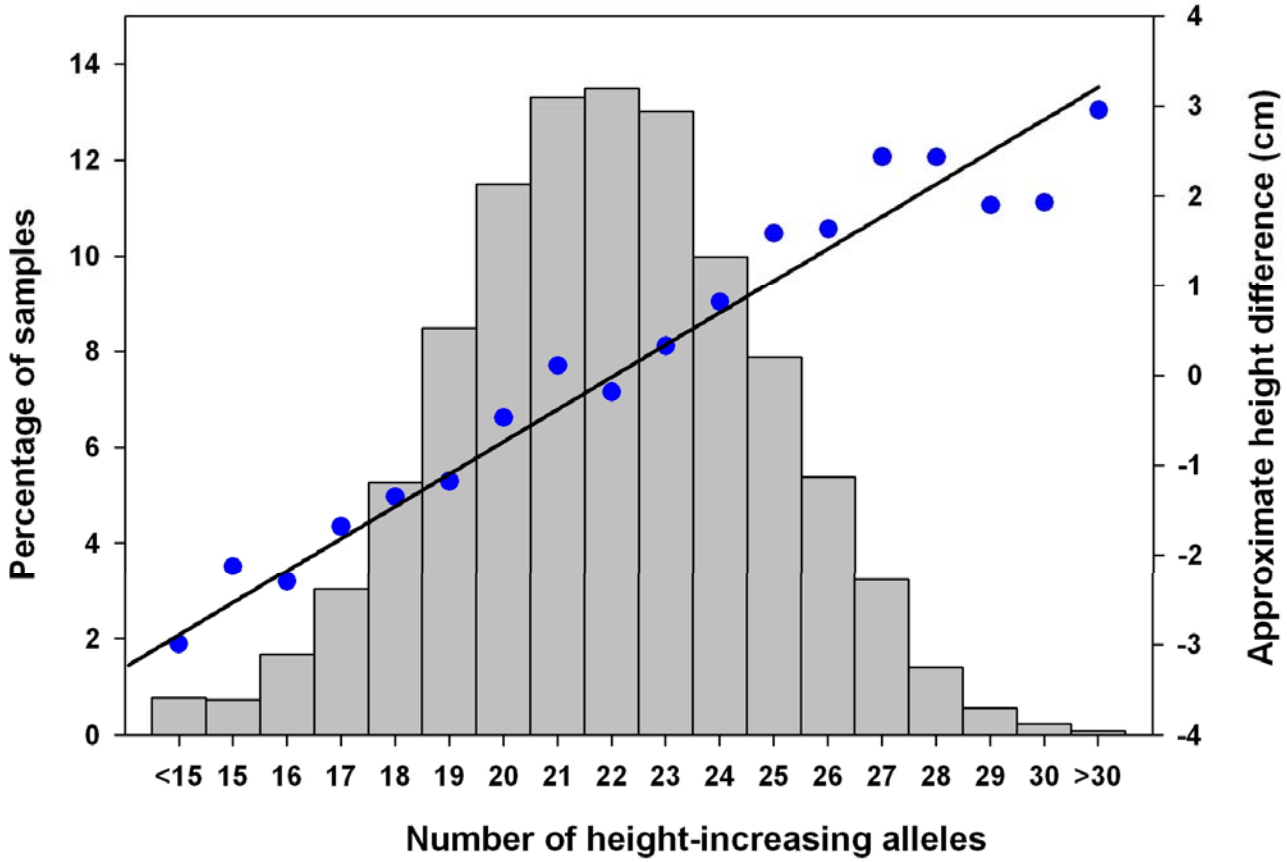
- Some real hits sit in the bottom of the distribution
- Some hits initially look interesting but then go away

Hedgehog signaling, cell cycle, and extra-cellular matrix genes over-represented

Candidate gene	Monogenic	Knockout mouse	Details*
ZBTB38	-	-	Transcription factor.
CDK6	-	Yes	Involved in the control of the cell cycle.
HMGA2	Yes	Yes	Chromatin architectural factors
GDF5	Yes	Yes	Involved in bone formation
LCORL	-	-	May act as transcription activator
LOC387103	-	-	Not known
EFEMP1	Yes	-	Extra-cellular matrix
C6orf106	-	-	Not known
PTCH1	Yes	Yes	Hedgehog signalling
SPAG17	-	-	Not known
SOCS2	-	Yes	Regulates cytokine signal transduction
HHIP	-	-	Hedgehog signaling
ZNF678	-	-	Transcription factor
DLEU7	-	-	Not known
SCMH1	-	Yes	Polycomb protein
ADAMTSL3	-	-	Extra-cellular matrix
IHH	Yes	Yes	Hedgehog signaling
ANAPC13	-	-	Cell cycle
ACAN	Yes	Yes	Extra-cellular matrix
DYM	Yes	-	Not known

Weedon et al. (in press) *Nat Genet*

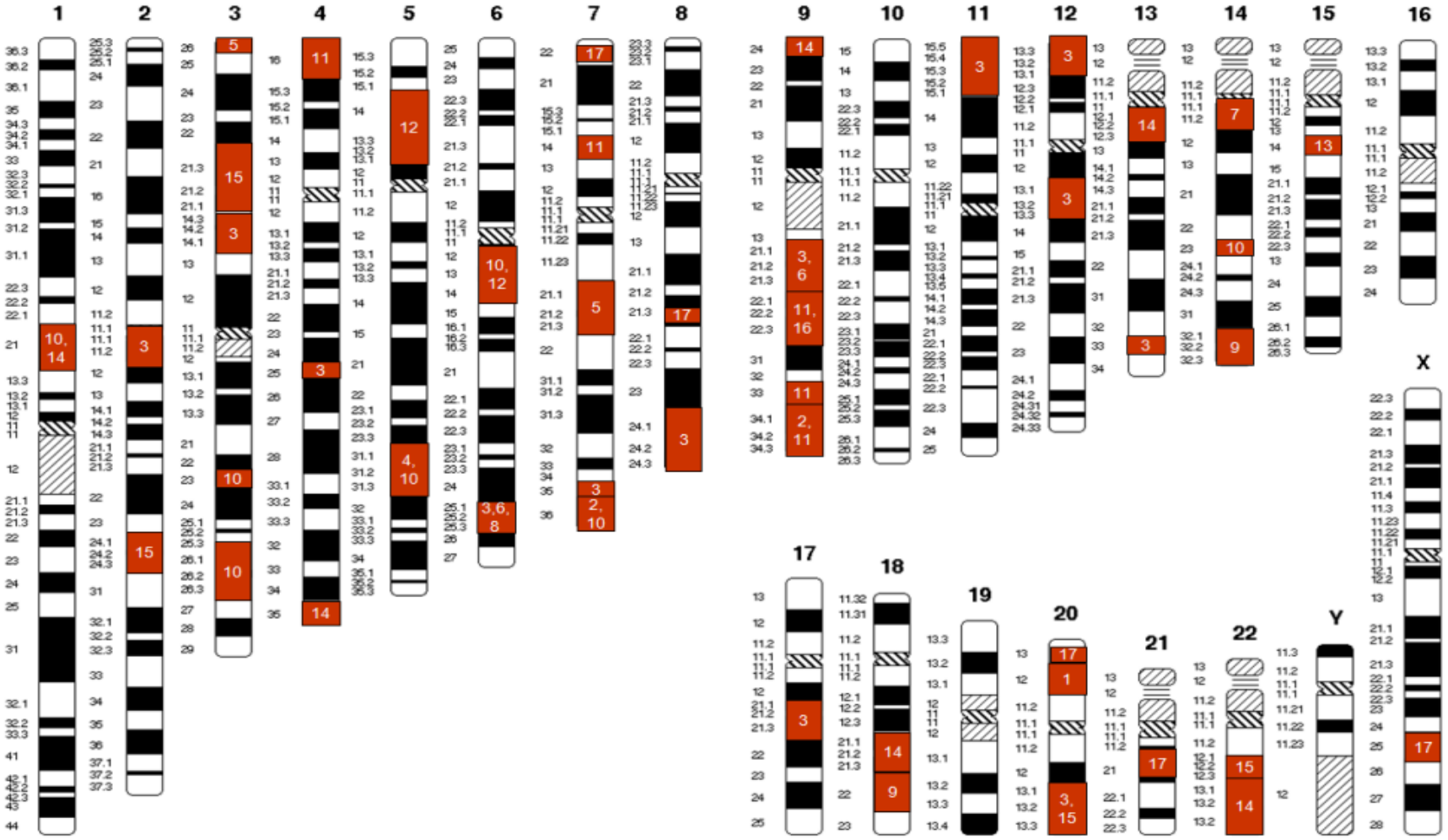
The combined impact of the 20 SNPS with a $P < 5 \times 10^{-7}$



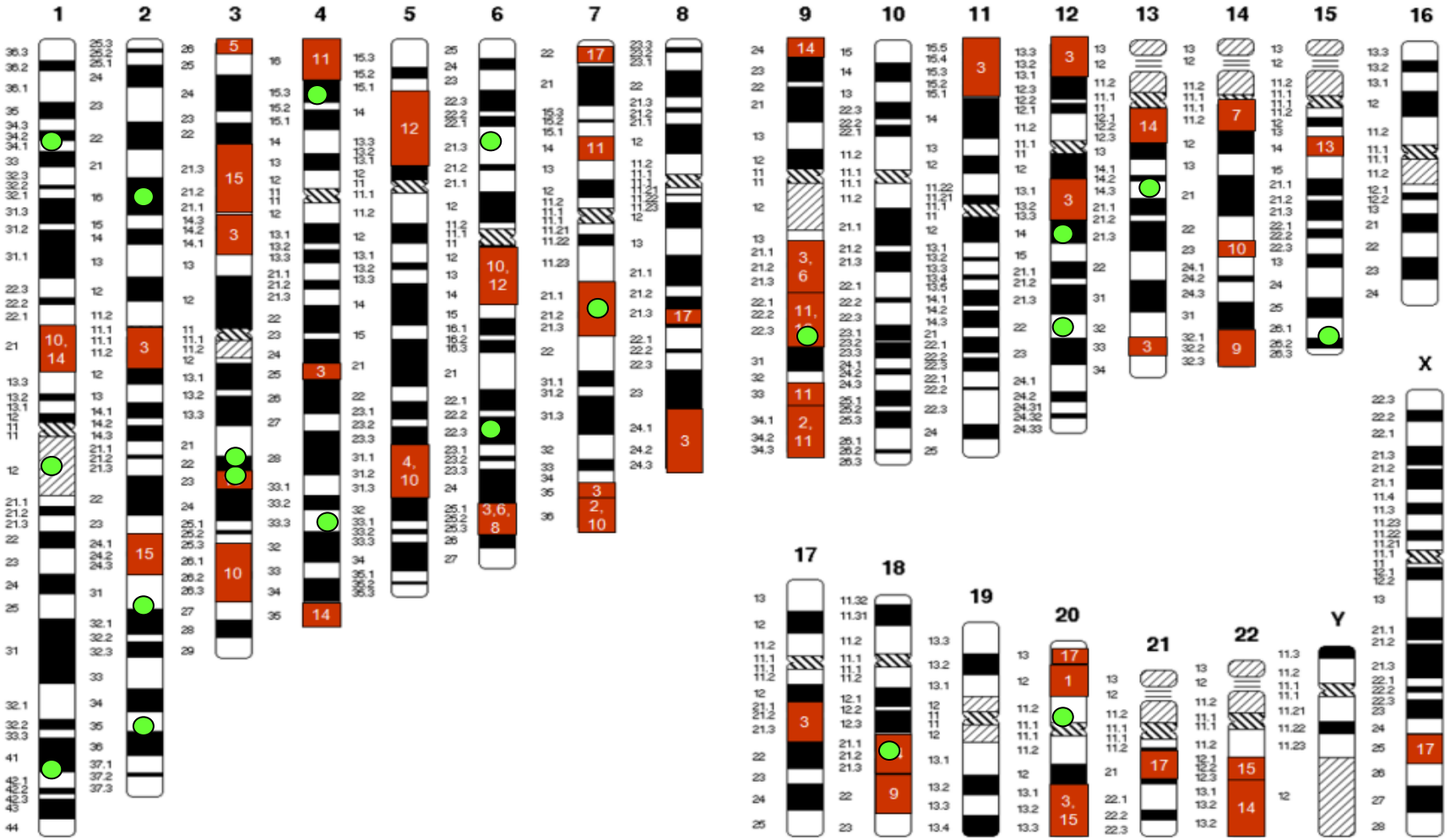
- The 20 SNPs explain only ~3% of the variation of height
- Lots more genes to find – but extremely large numbers needed

Weedon et al. (in press) *Nat Genet*

Height Linkage Regions



Perola et al, Plos Genetics, 2007; data available at http://www.genomeutwin.org; Weedon et al.; unpublished data



Perola et al, Plos Genetics, 2007; data available at http://www.genomeutwin.org; Weedon et al.; unpublished data

What's Going On?

- Loci identified by GWAs don't have linkage peaks over them

  Linkage analysis lacks power?

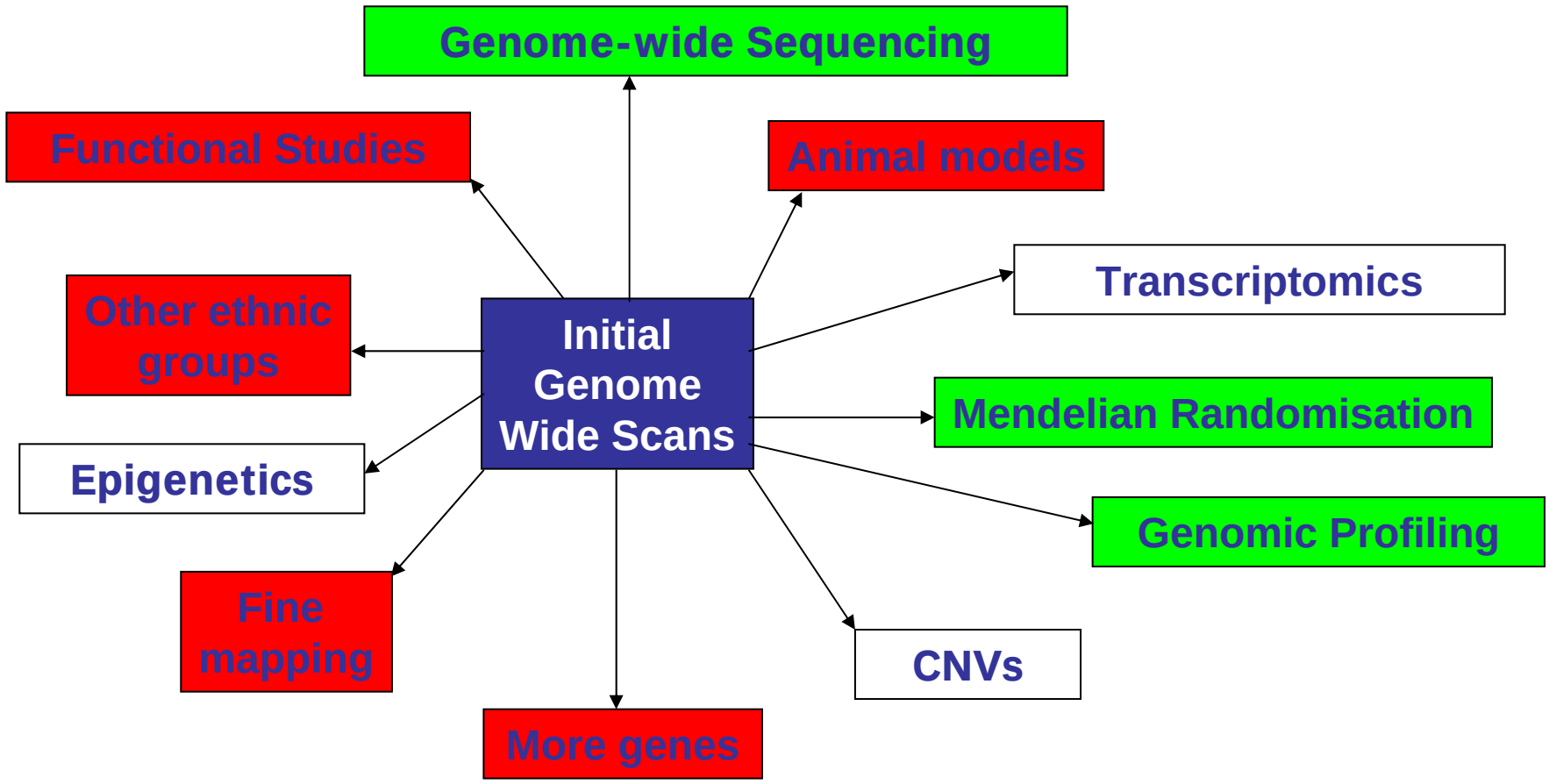
- Areas identified by linkage don't have significant assocation hits over them

  Type I error?

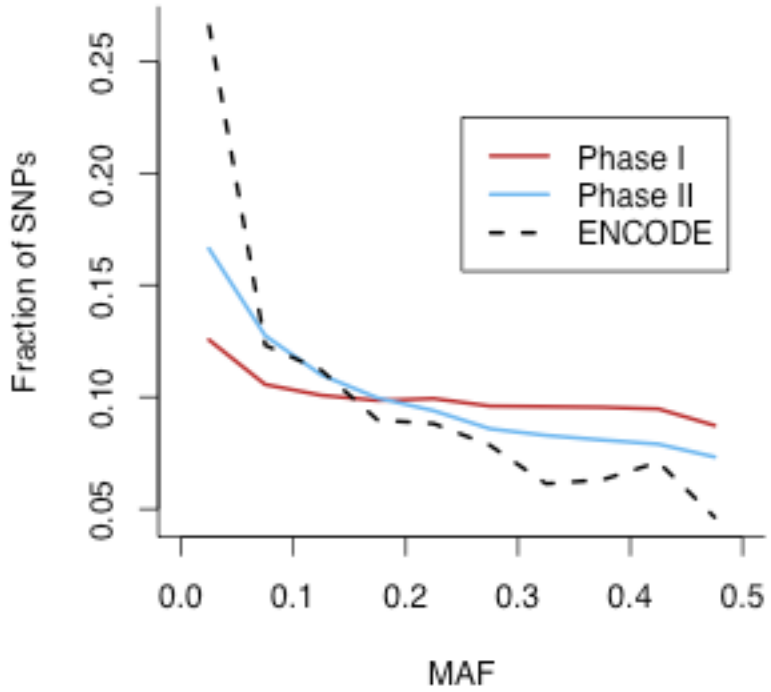
  Power?

- BUT…what if linkage analysis and association analysis identify different types of loci?

What next?

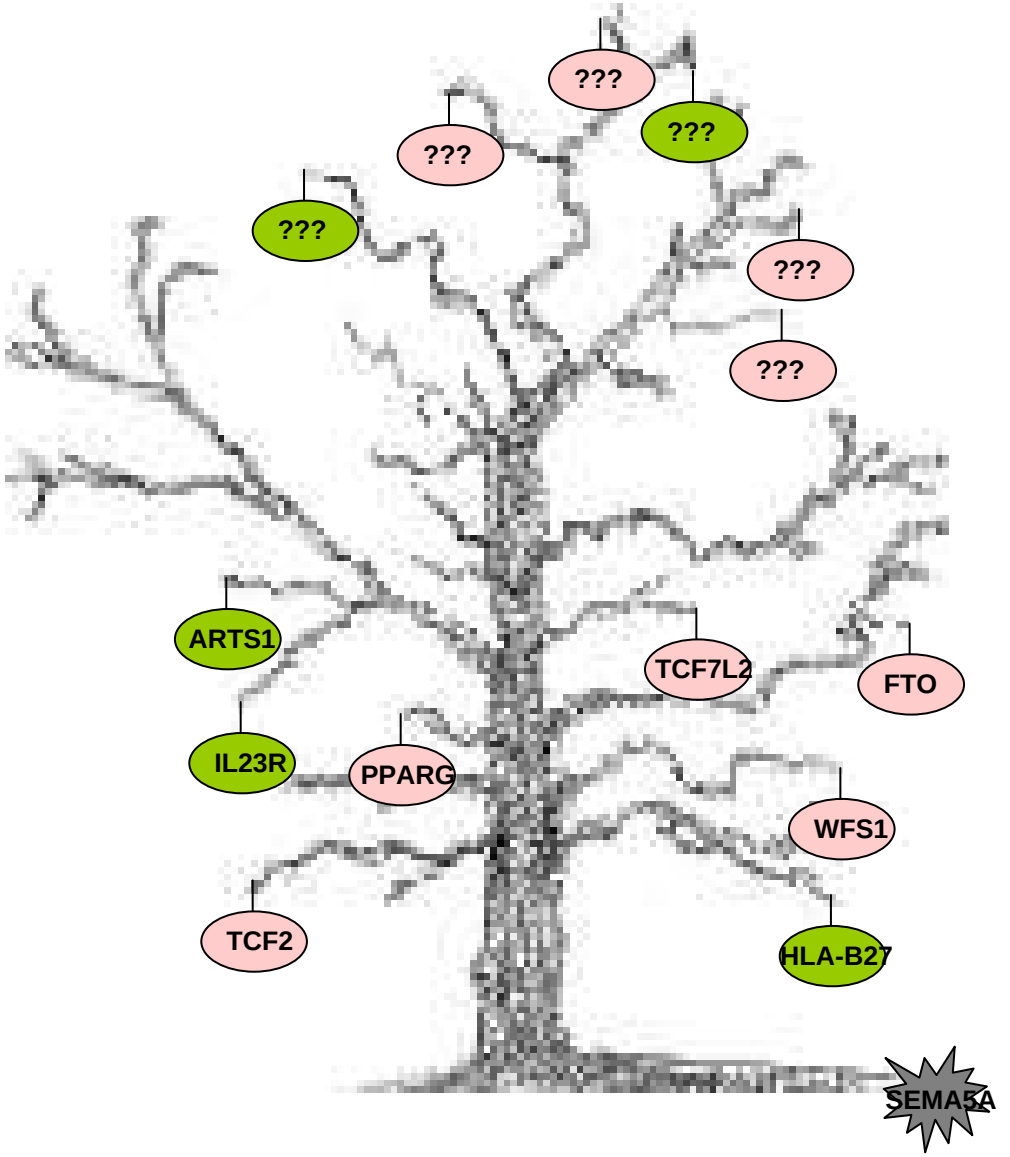

Genome-wide Sequencing
Functional Studies
Animal models
Transcriptomics
Other ethnic groups
Initial Genome Wide Scans
Mendelian Randomisation
Epigenetics
Genomic Profiling
Fine mapping
CNVs
More genes

Distribution of MAFs in HapMap



- Genome-wide panels and HapMap biased towards common variants
- Common variants don't tag rare variants well

Complex Disease Tree

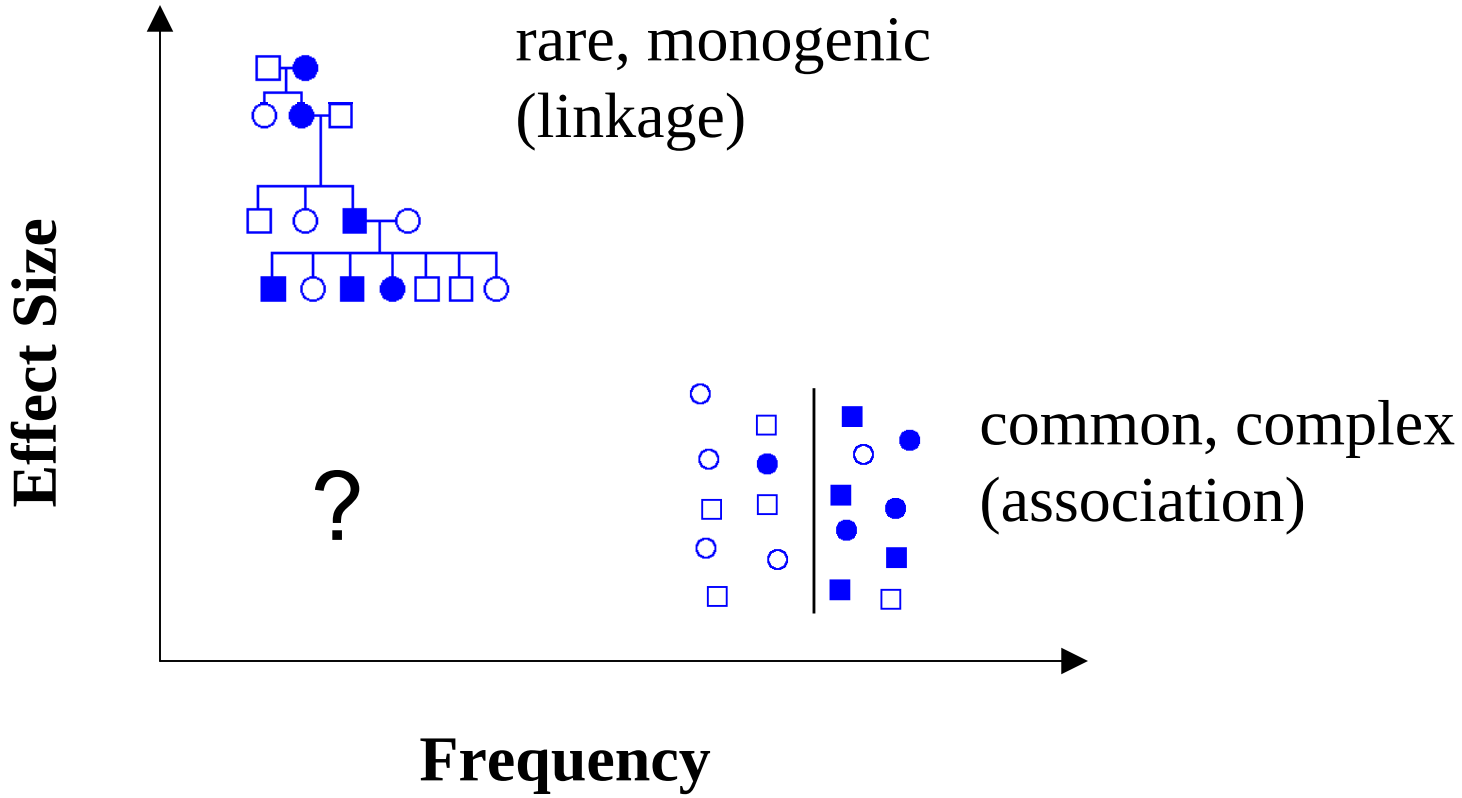


"High hanging fruit"

"Low hanging fruit"

"Rotten Fruit"

Methods of gene hunting



Genome-wide Sequencing

- Sequence individuals' genomes
- Will identify rare variants
- But will we have enough power?

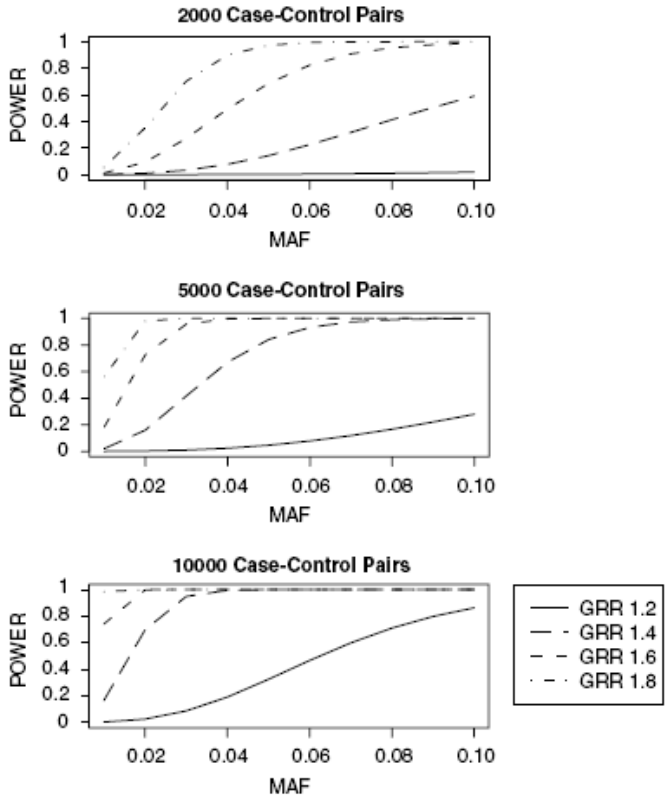


Figure 2 Relationship between MAF, heterozygote GRR, and power to detect association assuming a multiplicative disease model. Results are shown for 2000, 5000, and 10 000 case–control pairs assuming a disease prevalence of 1% and a type I error rate of $\alpha = 3.6 \times 10^{-6}$. The figure illustrates that it is possible to detect rare variants of intermediate penetrance using current sample sizes of 2000 case–control pairs. To detect rare alleles of smaller effect, far larger sample sizes will need to be employed.

Evans et al. (2008) *EJHG*

Genomic Profiling

- The idea of using genetic information to inform diagnosis
- Predictive testing in the case of monogenic diseases has been used for years (1300+ tests available) (e.g. Phenylketonuria)
- Not possible in complex diseases as effects of an individual variant is so small
- BUT…if we consider several predisposing genetic and environmental factors, can we predict disease?

Genomic Profiling

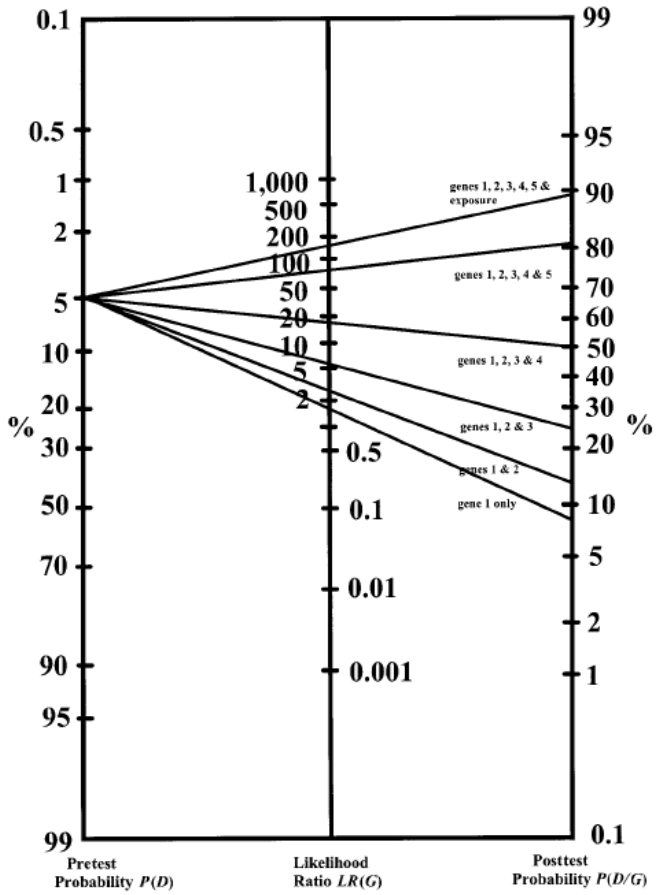
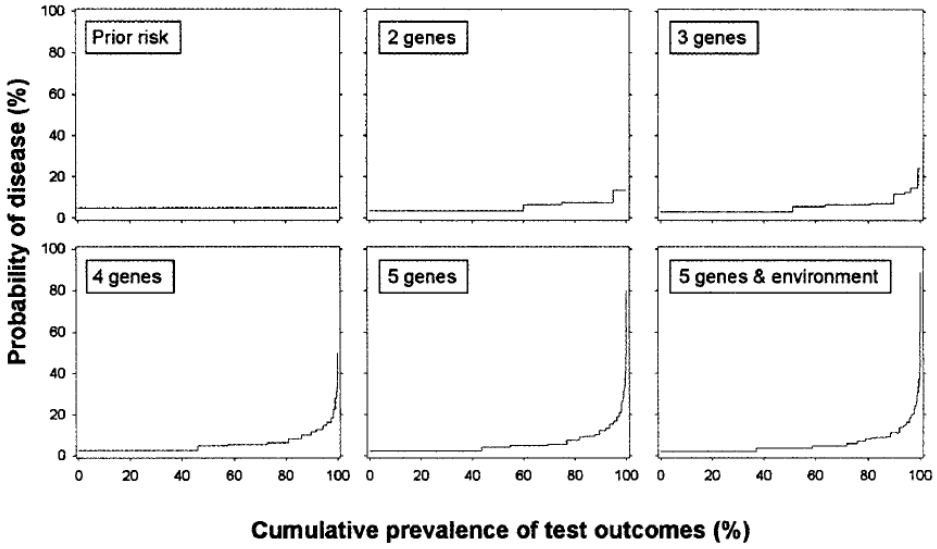


Figure 1 Power of a panel of genetic tests and exposure on predictability of the common disease (simulated data)

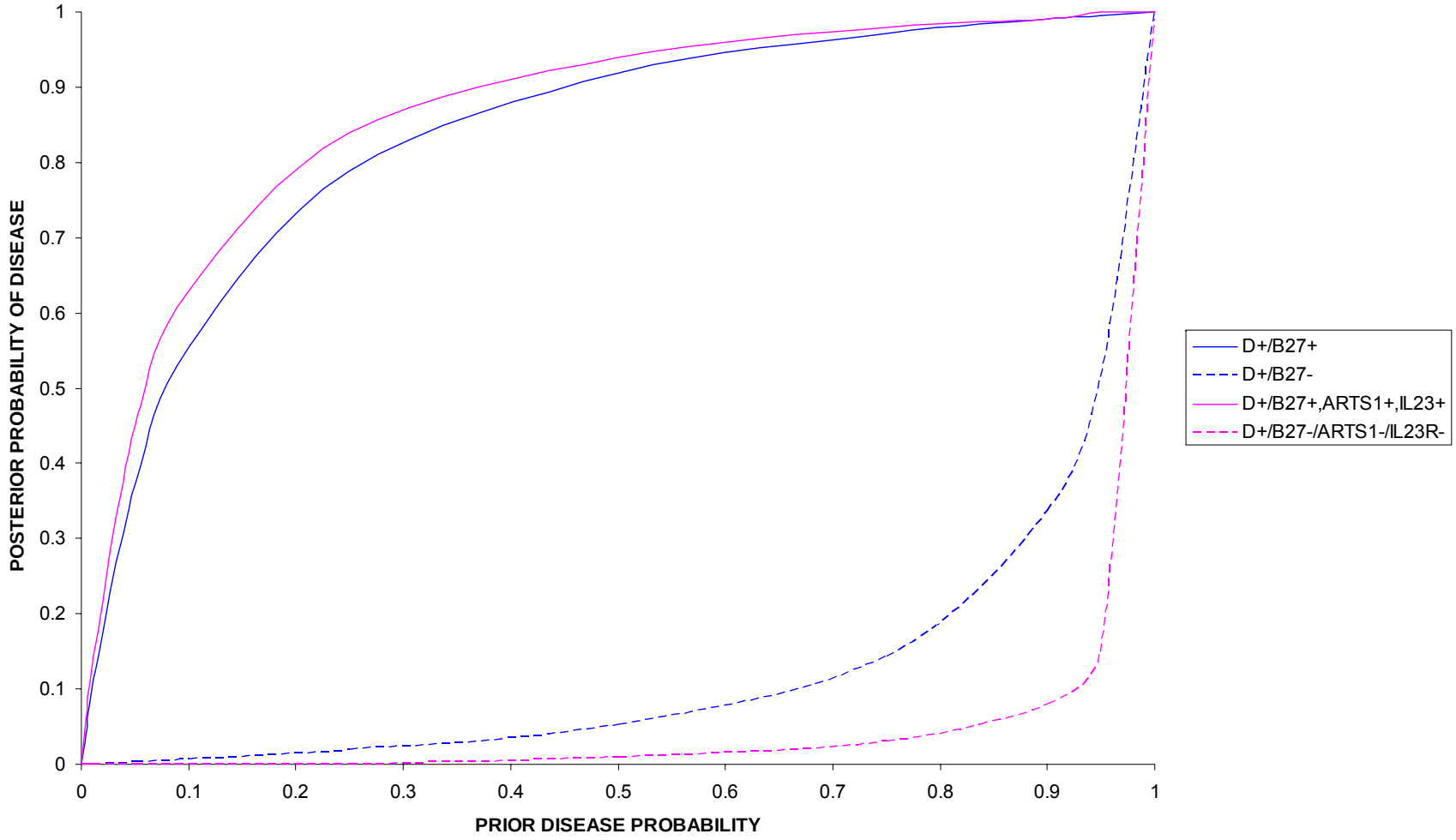
(from Yang et al. 2003 AJHG)



(from Janssens et al. 2004 AJHG)

=> Give up and go home?

Ankylosing Spondylitis



(Brown & Evans, in prep)

- Prevalence of B27+, ARTS1+,IL23R+ is 2.4%
- Prevalence of B27-, ARTS1-, IL23R- is 19%

Using Genetics to Inform Classical Epidemiology

Observational Studies

- Fanciful claims often made from observational studies
- In a case-control study, a group of diseased individuals are recruited (Cases); A group of individuals without disease are gathered (Controls); Both groups are then measured retrospectively on an exposure of interest; A test of association is performed
- Example: Obesity (Exposure) and Coronary Heart Disease (Outcome)

	Obesity	
	Yes	No
CHD	200	100
Control	50	250

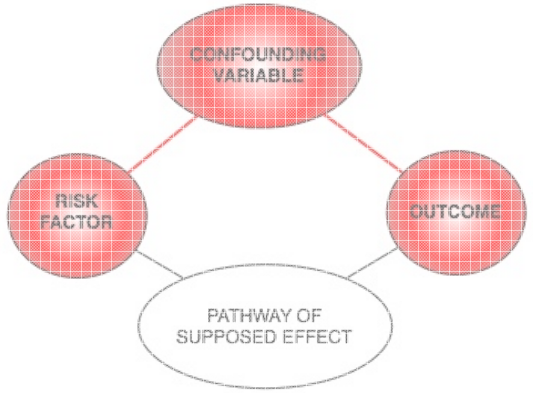
Odds of obesity in cases: 200/100 = 2

Odds of obesity in controls: 50/250 = 0.2

Odds Ratio: 2/0.2 = 10

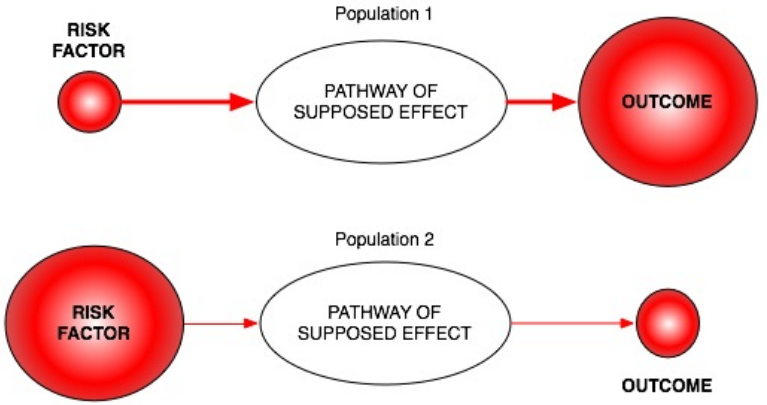
Classic limitations to "observational" science

- Confounding

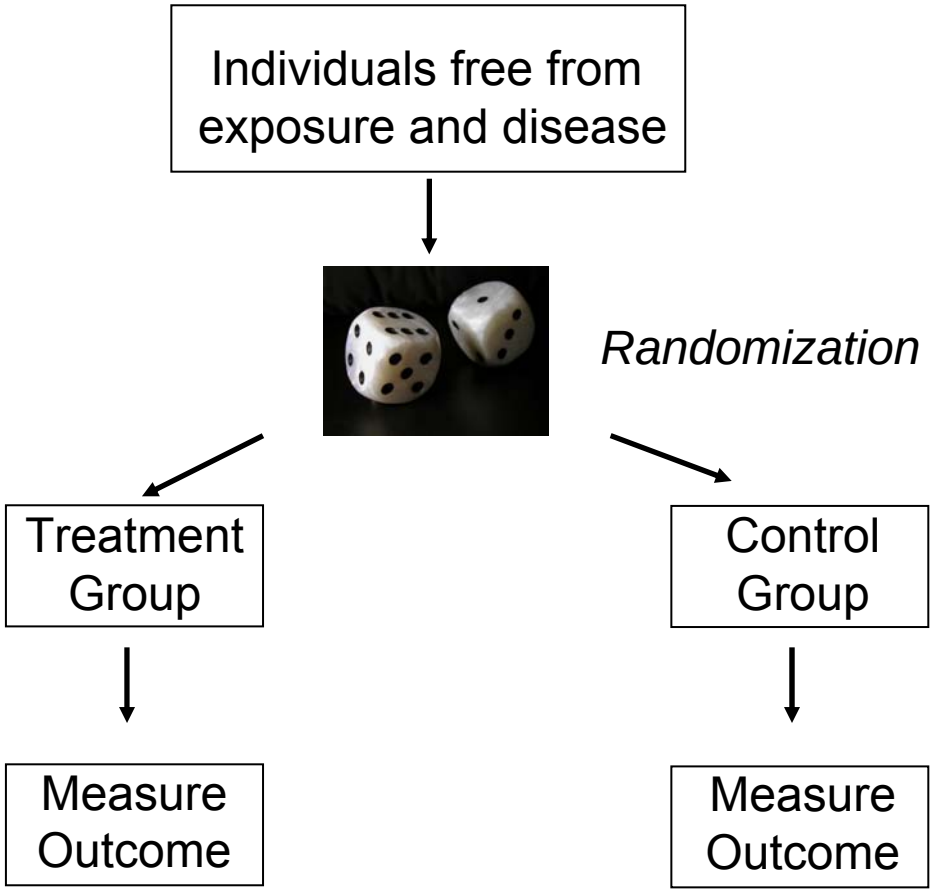


- Reverse Causation

- Bias



Randomized Control Trials

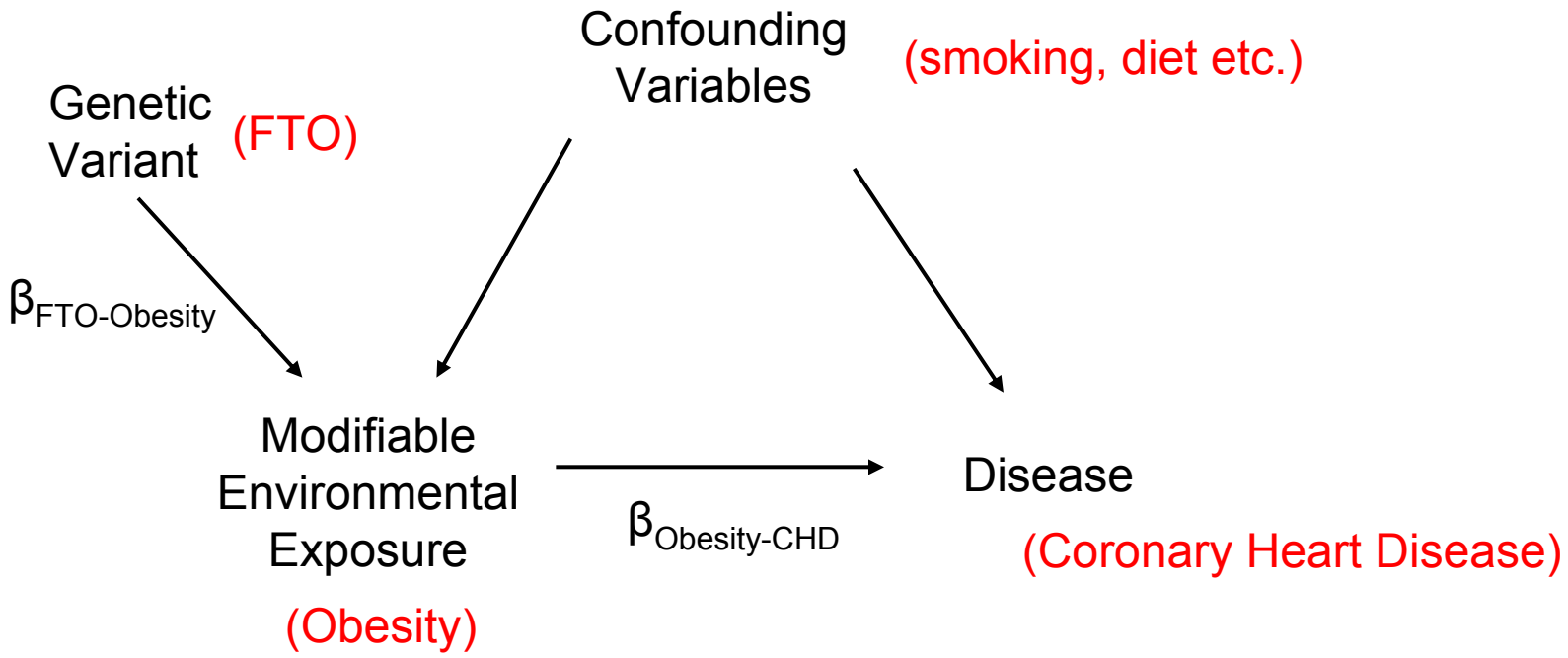


- Randomization controls for confounding
- Reverse causation impossible
- Gold standard for assessing causality

Mendelian Randomization

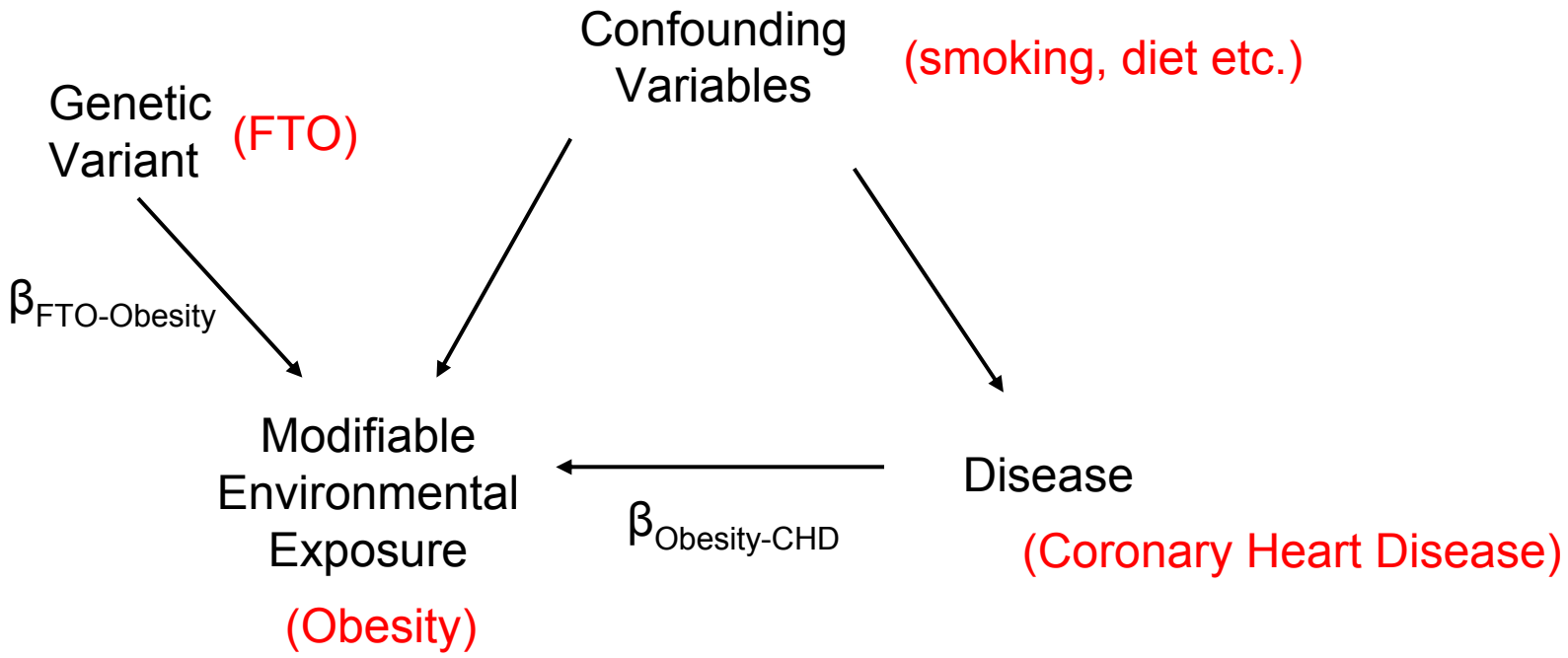
- RCTs not always ethical or possible
- Fortunately nature has provided us with a natural randomized control trial !
- Mendel's law of independent assortment states that inheritance of a trait is independent (randomized) with respect to other traits
- Therefore individuals are randomly assigned to three groups based on their genotype (AA, Aa, aa) independent of outcome
- Assessing the relationship between genotype, environmental risk factor and disease informs us on causality

Mendelian Randomization



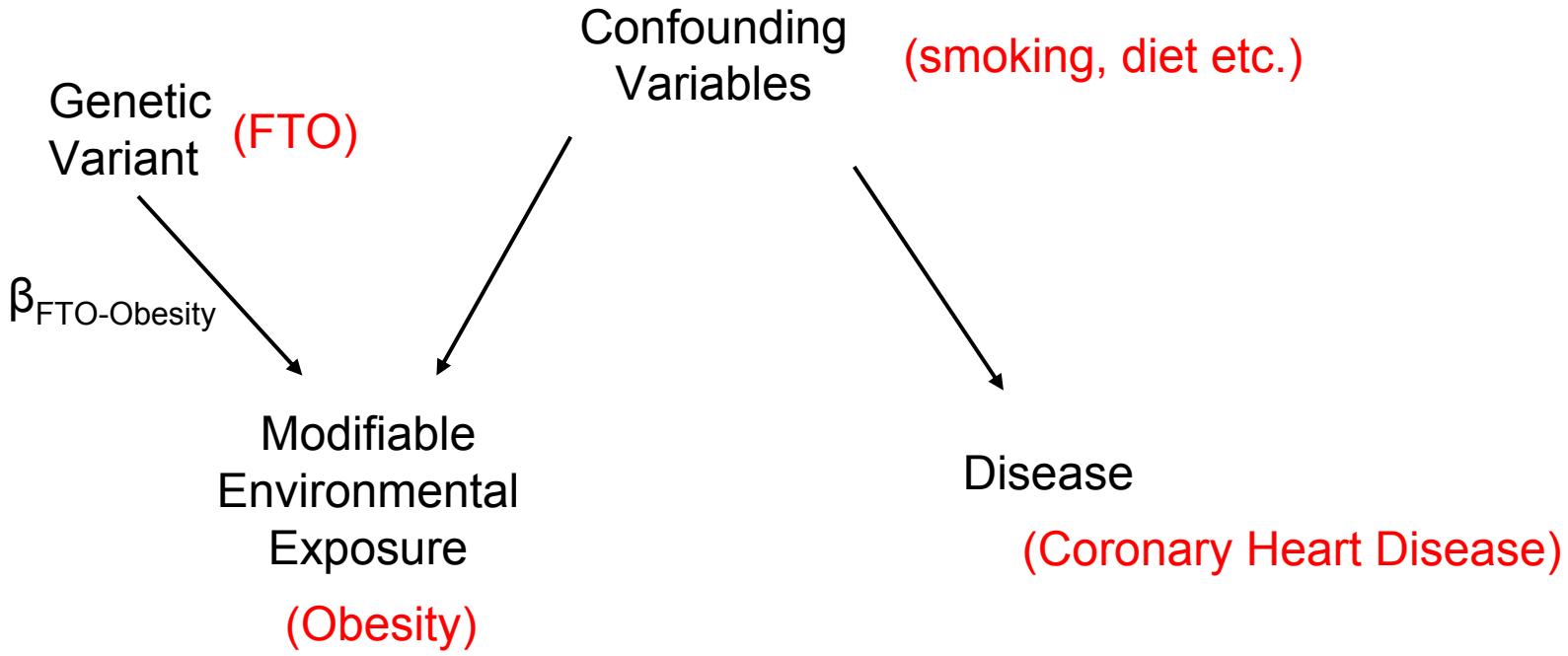
- If obesity causes CHD then the relationship between FTO and CHD should be estimated by the product of $\beta_{FTO\text{-}Obesity}$ and $\beta_{Obesity\text{-}CHD}$

Mendelian Randomization



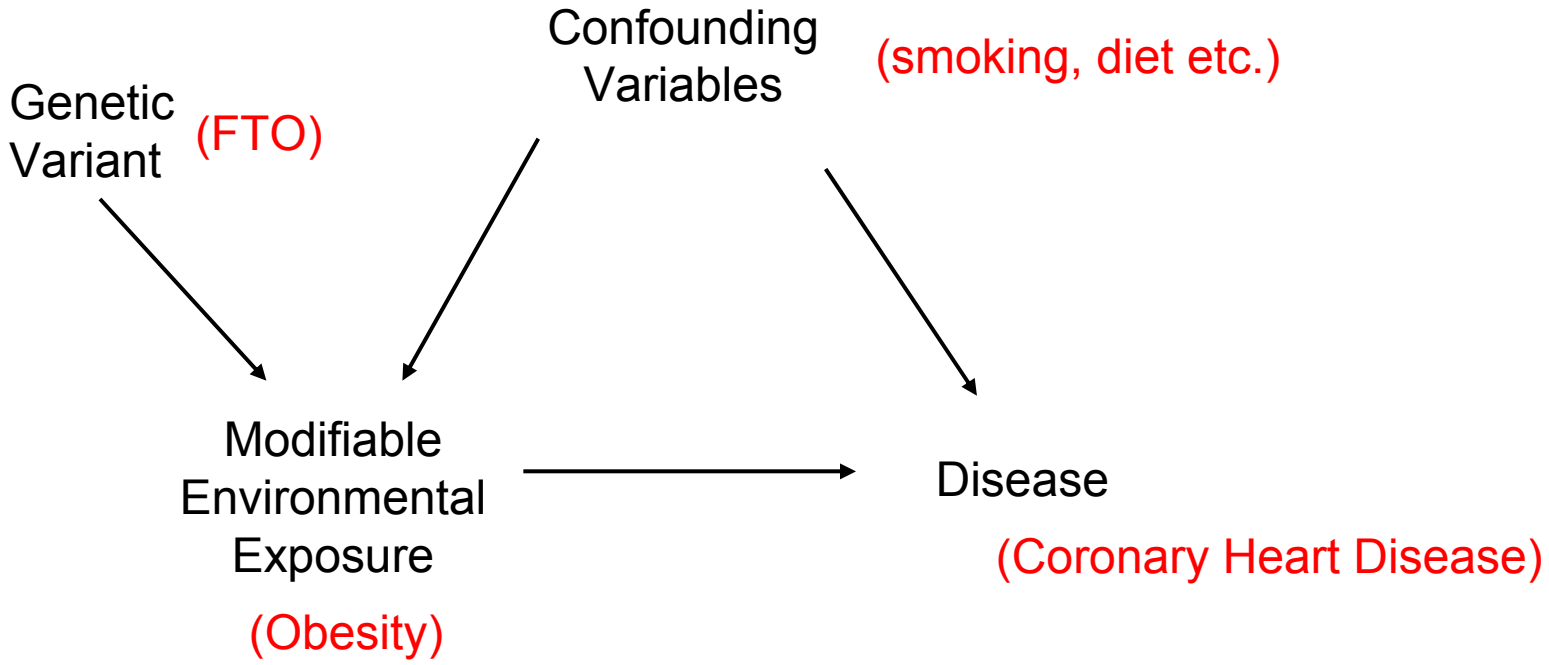
- If CHD causes obesity then $\beta_{FTO\text{-}CHD}$ should be zero.

Mendelian Randomization



- If the relationship between Obesity and CHD is purely correlational (i.e. due to confounding) then $\beta_{FTO\text{-}CHD}$ should be 0

Mendelian Randomization



- Genotype is associated with the environmental exposure of interest
- Genotype is NOT associated with confounders
- Genotype is only related to its outcome via its association with the modifiable environmental exposure

Mendelian Randomization

- Mendelian Randomization is a way of using a genetic variant(s) to make causal inferences about (modifiable) environmental risk factors for disease and health related outcomes

- Environmental exposures (e.g. Obesity) can be modified ! Genetic factors cannot (at least for the moment…)

- Still a relatively new approach that has problems (i.e. finding genetic proxies for environmental exposures- multiple instruments?)

…but a LOT of scope for development…

Could SEM be used to enhance MR?